DOCUMENT RESUME

ED 322 855 HE 023 803

AUTHOR Banta, Trudy W.

TITLE Assessment Resource Center, 1986-1989. Final Report.

INSTITUTION Tennessee Univ., Knoxville. Center for Assessment Research and Development.

SPONS AGENCY Fund for the Improvement of Postsecondary Education (ED), Washington, DC.

PUB DATE 90

CONTRACT DEG-G008642196

NOTE 143p.

PUB TYPE Reports - Descriptive (141) -- Reference Materials - Bibliographies (131)

EDRS PRICE MF01/PC06 Plus Postage.

DESCRIPTORS *College Outcomes Assessment; *Evaluation Methods; Higher Education; Instructional Effectiveness; *Measurement Techniques; Outcomes of Education; *Program Implementation; *Student Evaluation

IDENTIFIERS Assessment Resource Center TN

ABSTRACT

Since its establishment with a 1986 grant from the Fund for the Improvement of Postsecondary Education, the Assessment Resource Center (ARC) at the University of Tennessee has worked with state coordinating boards, other institutions involved in assessment leadership, institutions seeking guidance, and participants in conferences and seminars given by the ARC. The Center accomplished its objectives of establishing working relationships with numerous institutions, preparing printed materials, sponsoring workshops, and developing a consortium of experienced assessment practitioners. In addition, Center personnel founded a national publication, planned an international seminar, and supported a cross-national study of assessment in higher education. This final report outlines the project's background and purpose, summarizes project impact, and documents plans for continuation and dissemination. Appendices, which comprise the bulk of the report, include: (1) an overview of the ARC; (2) an annotated bibliography of 12 items on assessment and a list of 23 representative assessment programs; (3) research on the College Outcome Measures Project, with an 11-item annotated bibliography; (4) title pages and tables of contents of ARC publications; (5) resource appendices from "Performance and Judgment: Essays on Principles and Practice in the Assessment of College Student Learning" edited by Clifford Adelman, containing an annotated bibliography of approximately 75 items and reviews of 22 assessment instruments. (JDD)

****

* Reproductions supplied by EDRS are the best that can be made *

* from the original document. *

****

HE

ED322855

ASSESSMENT RESOURCE CENTER

1986-1989

FINAL REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

A Project Supported
by the
Fund for the Improvement
of Postsecondary Education

and the

Center for Assessment Research and Development
1819 Andy Holt Avenue
Knoxville, Tennessee 37996-4350
(615) 974-2350

HE023803

BEST COPY AVAILABLE

COVER SHEET

Grantee Organization:

University of Tennessee, Knoxville
Learning Research Center (Now the Center for Assessment Research & Development)
1819 Andy Holt Avenue
Knoxville, TN 37996-4350

Grant No.: DEG G008642196

Starting Date: September 15, 1986
Ending Date: September 14, 1989
Number of months: 36

Project Director:

Trudy W. Banta
Center for Assessment Research & Development
University of Tennessee, Knoxville
1819 Andy Holt Avenue
Knoxville, TN 37996-4350

FIPSE Program Officer(s): Tom Carroll and David Holmes

Grant Award:

Year 1	$100,643.00
Year 2	125,046.00
Year 3	147,990.00
Total	373,679.00

Assessment Resource Center Project

Table of Contents

Abstract....................ii
Executive Summary iii
Overview, Background, and Purpose 1
Project Description 4
Results....................12
Plans for Continuation and Dissemination 14
Summary and Conclusions 15
Appendixes....................18

ABSTRACT

Since its establishment with a FIPSE grant in 1986, the Assessment Resource Center (ARC) has worked with state coordinating boards, other institutions involved in assessment leadership, institutions seeking guidance from the center, and several hundred participants in conferences and seminars given by the Center. The Center accomplished all of its original objectives, establishing working relationships with numerous institutions, preparing printed materials, sponsoring workshops, and developing a consortium of experienced assessment practitioners. In addition, Center personnel founded a national publication, planned an international seminar, and supported a cross-national study of assessment in higher education. We have worked to meet the increasing demand for materials and assistance, and hope to continue to contribute to the assessment literature and to participate in ongoing discussions about assessment at national and international levels.

Trudy W. Banta
Center for Assessment Research and Development
1819 Andy Holt Avenue
Knoxville, Tennessee 37996-4350
(615) 974-2350

Assessment Resource Center Publications

Articles on Assessment at UTK
Assessment Bibliography
Bibliography of Assessment Instruments
Material on Locally-Developed Tests
Satisfaction Surveys Used at UTK
Research Papers on Assessment
1988 Strategies for Assessing Outcomes Handouts
Assessment of Student Outcomes in Higher Education - Research Reports by UTK Faculty

ASSESSMENT RESOURCE CENTER PROJECT

Assessment Resource Center
1819 Andy Holt Avenue
Knoxville, Tennessee 37996-4350

Trudy W. Banta, Project Director
Gary R. Pike, Associate Director

EXECUTIVE SUMMARY

Project Overview

The national Assessment Resource Center project began at the University of Tennessee, Knoxville in September 1986. Outcomes assessment had become a priority in higher education as a result of institutional responses to recommendations in several national reports and in actions taken by a number of states and six regional accrediting boards. The state of Tennessee had become a leader in requiring public institutions to report the results of outcomes assessment activities. The University of Tennessee, Knoxville (UTK) had achieved national recognition for its assessment program and was receiving requests for assistance from other institutions across the nation who were beginning to develop such programs of their own. Thus representatives of the University proposed that FIPSE provide support to establish a national Assessment Resource Center at UTK.

Since its establishment with a FIPSE grant in 1986, the Assessment Resource Center (ARC) has worked with state coordinating boards, other institutions involved in assessment leadership, institutions seeking guidance from the center, and several hundred participants in conferences and seminars given by the Center. The Center accomplished all of its original objectives, establishing working relationships with numerous institutions, preparing printed materials, sponsoring workshops, and developing a consortium of experienced assessment practitioners. In addition, Center personnel founded a national publication, planned an international seminar, and supported a cross-national study of assessment in higher education.

Background and Purpose

In 1984 educational leaders from UTK and several other institutions with successful assessment programs began meeting in small groups sponsored by the Association of American Colleges (AAC) and the American Association of Higher Education (AAHE) to discuss the need for national leadership for assessment. As a research university with a nationally and internationally recognized comprehensive outcomes assessment program, UTK had the resources necessary to establish and maintain a national center to provide information and coordinate developmental efforts on the topic of assessment. Since its establishment at UTK, the ARC has provided services to hundreds of participants through campus consultations, conference papers, publications, a national newsletter, on-site conferences, and an international conference.

During the period of its grant, the ARC has operated with a part-time director, a full-time associate director, a graduate assistant, and a secretary, at a state institution with an enrollment of 25,000 undergraduate and graduate students. Interest in assessment has increased over the life of the project, and the need for the services originally proposed has grown annually both in terms of the numbers of new learners to orient and new topics to discuss. We have learned that there is a greater need for what we proposed than we could ever have imagined, and our redefinition of the project has been to extend our services to a larger and more varied audience.

Project Description

The ARC proposal included four objectives: (1) to develop a bibliography of instruments and practices; (2) to gain recognition as a national resource center; (3) to provide information to others; and (4) to be active in discussing and offering solutions for assessment issues. For the ARC to accomplish these objectives we realized that the bibliography should be iterative, the center would have to gain visibility prior to establishing leadership, information would best be provided in person to participants beginning to establish assessment programs, and small forums would best unite individuals focused on specific issues.

Developing an assessment instrument bibliography became the first priority of the Center because that document was needed for distribution at the first workshop. Separate bibliographies on practices then were planned. A related project funded by the Office of Educational Research and Improvement enabled us to compile and publish one of the latter bibliographies.

Frequent presentations at national and regional meetings became the principal means of gaining visibility for the ARC, and smaller conferences, consultations, workshops, and campus visits furthered achievement of this objective. Publishing important papers helped establish the ARC as a leader in assessment, and during the first year we began to make our printed materials available by mail upon request.

Project Results

Services in the form of materials and consultations have been provided to state agencies and colleges and universities in 49 states, the District of Columbia, Puerto Rico, and 5 foreign countries. Presentations have been made at the annual meetings of the AAC, AAHE, AEA, AERA, AIR, and FIPSE project directors. Three workshops were held in 1987 and 1988, with a total of more than 500 people attending. With help from the FIPSE project grant, the ARC published 27 papers, reports, and books; made 85 invited presentations off campus; developed a consortium of campus-based leaders; assumed a leadership role in the assessment movement; and cooperated with the AAHE Assessment Forum, the leading higher education organizations, and several state higher education agencies. In 1988 Jossey-Bass Publishers, Inc. invited the project director to edit the first national quarterly on assessment in higher education, Assessment Update, which reaches 1800 subscribers.

Over the three years of this project, a change occurred in the type of information requested, moving from the most basic how-to-get-started inquiries to questions about more substantive methodological issues, such as the best way to measure achievement. Also, there has been growing interest in such non-traditional measures as portfolios and self-assessment. Presentations have changed from descriptions of UTK's assessment program to discussions of a variety of methodologies for assessing student outcomes. Contacts are handled differently now, by mail and telephone, with printed materials being sent out, rather than by receiving visitors at UTK. Now faculty as well as administrators in institutions of higher education, including UTK, seek to improve learning for their students as they respond to the various external pressures to become involved in assessment.

The ARC gauged the impact of the project by the amount of information distributed and the quality of that information as rated by users via written evaluations of printed materials and workshops. Respondents to mailed surveys provided a mean overall rating of ARC materials of 4.1 on a 5-point scale ranging from poor (1) to excellent (5). Three workshops were consistently given ratings in excess of 7 points on a 10-point scale. Presentations, articles, and campus visits increased, and 8 major reference works, 19 research reports, and an international conference and study-group report were produced. The four objectives were accomplished, and the national interest in assessment continues undiminished.

Continuation and Dissemination

Plans for continuation include maintaining our leadership position with AAC, AEA, AAHE, AERA, and AIR presentations this year, as well as papers, symposia, and workshops. Evaluation activities to be continued include monitoring the quality of our materials via user surveys. We also plan to assess the quality of future workshops using printed evaluation forms similar to those developed in the course of the ARC project. In a major step following completion of the project, the University of Tennessee established the Center for Assessment Research and Development (CARD) to continue the work of the ARC. New projects for CARD include development of an employer survey and a second international conference, this one at St. Andrews University in Scotland in July 1990.

Conclusions

Insights gained as a result of grant activity stem from our awareness of the public's growing interest in accountability and favorable reactions to workshops and to assessment materials that we have made available. We have worked to meet the increasing demand for materials and assistance, and hope to continue to contribute to the assessment literature and to participate in ongoing discussions about assessment at national and international levels.

OVERVIEW, BACKGROUND, AND PURPOSE

In 1986, when this project was proposed, there were clear indications that outcomes assessment in higher education was becoming an issue of high priority for college and university administrators and faculty. The following items illustrate this point:

- Several national reports, notably the National Institute of Education's (NIE) Involvement in Learning and Integrity in the College Curriculum from the Association of American Colleges (AAC), had called for institutions to assess the outcomes of student learning.
- A national conference planned to investigate response to the NIE report one year after its release in 1984 became the first National Conference on Assessment in Higher Education and attracted over 700 people -- nearly double the early estimates of attendance.
- Two task forces of the National Governors' Conference had recommended that colleges and universities begin to furnish concrete evidence of their accountability for student intellectual development.
- The six regional accrediting associations had revised their standards to require evidence of student accomplishment of the educational objectives set by each institution.
- In Tennessee and South Dakota state coordinating boards had required public institutions to administer certain standardized tests and report students' scores to the respective boards. In other states, such as New Jersey and Virginia, coordinating agencies had supported pilot projects in selected institutions to encourage public colleges and universities to plan individualized outcomes assessment programs.

In 1986 only three institutions -- Alverno College, Northeast Missouri State University and the University of Tennessee, Knoxville (UTK) -- had acquired national recognition for their comprehensive outcomes assessment programs. Small groups of educational leaders convened in 1985 by the AAC and the American Association for Higher Education (AAHE) had expressed concern that the growing number of institutions interested in outcomes assessment had no central resource to consult for assistance. Faculty and administrators at Alverno, Northeast Missouri and UTK were becoming overwhelmed by calls for help from other institutions.

UTK proposed that FIPSE support an Assessment Resource Center (ARC) at a research university for the purpose of addressing the burgeoning need for information, assistance, and leadership in the area of outcomes assessment in higher education. To quote the original project proposal, "As the research university with a nationally and internationally recognized comprehensive outcomes assessment program, The University of Tennessee, Knoxville is uniquely qualified to . . . develop and disseminate information, address concerns common to groups of institutions, and improve practice in the assessment of student outcomes in higher education." Specific objectives of the ARC were to:

(1) Develop at least one annotated bibliography of assessment instruments and practices for distribution to other institutions.

(2) Gain recognition throughout the academic community for the ARC.

(3) Provide information about assessment materials and procedures to those seeking such information.

(4) Contribute to the discussion and solution of assessment issues and problems by bringing together appropriate individuals for conversation and action.

With excellent support from the faculty and administrative staff of the University of Tennessee and of the FIPSE staff, all of the original objectives of the Assessment Resource Center have been accomplished. The staff, consisting of a half-time director, a full-time associate director, a graduate assistant, and a secretary, has developed and disseminated a continuously updated annotated bibliography of assessment instruments. The ARC, now the Center for Assessment Research and Development, is sufficiently well-known to have provided information and/or assistance to colleges and universities in 49 states, as well as institutions and coordinating agencies in a half-dozen foreign countries.

Excellent working relationships have been established with virtually all of the other institutions that have received early recognition as assessment leaders: Alverno, Northeast Missouri, Miami-Dade Community College, James Madison University, Kean College of New Jersey, King's College of Pennsylvania. Substantial amounts of staff time have been spent working with the state coordinating agencies of Tennessee, New Jersey and Virginia and with individual institutions in these states, as well as those in Missouri, California, Georgia, Pennsylvania, and South Carolina. In total, campus visits have been arranged to provide assistance to 52 institutions in 24 states and Puerto Rico.

Four workshops, attracting between 95 and 204 participants each, were conducted with FIPSE assistance between 1986 and 1989. In a related development, an international seminar at Cambridge University for 45 individuals from 9 countries was supported by the University of Tennessee in July 1989. Also, in 1989 Jossey-Bass Publishers of San Francisco selected the project director, Trudy Banta, to edit Assessment

Update, the first newsletter project ever undertaken by J-B and the first periodical devoted exclusively to assessment in higher education.

The project contributed to the solution of problems in the field through a series of original contributions to the research literature and the creation of a consortium of the field's most experienced practitioners that met periodically for discussion and action.

Following through on its initial commitment to support the ARC when FIPSE funds were withdrawn, on July 1, 1989 the University of Tennessee established the Center for Assessment Research and Development (CARD). This center continues to provide written materials and personal assistance, as well as annual workshops, for other institutions seeking advice about assessment. A second international seminar is scheduled at St. Andrews University in Scotland July 24-27, 1990. A new FIPSE grant will permit UTK to coordinate a consortium of Tennessee institutions planning to develop an employer survey to enhance their assessment efforts.

PROJECT DESCRIPTION

The goal of the Assessment Resource Center was to "enhance the quality of the educational experience for students by increasing the use of outcomes assessment procedures in evaluating the progress of individual students as well as the quality of programs and services within higher education". In order to attain this goal, four specific objectives were established to guide program implementation and serve as a basis for evaluation. A brief summary of the efforts made by Center staff to meet these objectives is presented in Appendix A.

Objective 1: Develop at least one annotated bibliography of assessment instruments and practices. In attempting to meet the first objective during 1986, it became apparent to the Center's staff that more than one bibliography would be required. Furthermore, it was decided that the development of these bibliographies would be an iterative process. Given the interest in standardized outcomes measures, and facing a March 1987 deadline for the Center's first assessment workshop, the project director and associate director decided that a Bibliography of Assessment Instruments should be the first priority. Using information from the ETS Test Collection Annotated Bibliography of Tests, this first bibliography outlined the basic characteristics of each test and provided addresses for further information. Participation in the Office of Educational Research and Improvement's assessment audit project during the Center's second year helped to expand the Bibliography of Assessment Instruments, particularly in the areas of the measurement of critical thinking and motivation. The OERI assessment audit also provided the impetus for a second bibliography concerning articles on assessment practice. Edited versions of both the Bibliography of Assessment Instruments and the articles on assessment practice are contained in the Performance and Judgment volume published by OERI.

The project director's editorship of a volume in the New Directions for Institutional Research series during the Center's second year provided a vehicle for developing a list of key resource people and institutions which was published in the New Directions volume. A copy of this chapter is presented in Appendix B. The materials developed in conjunction with the OERI assessment audit and the New Directions series were used to revise the Bibliography of Assessment Instruments and were presented to workshop participants during the second year of the project. To date, approximately 600 copies of the Bibliography of Assessment Instruments have been distributed.

During the third year of the project, editorial changes were made in the Bibliography of Assessment Instruments and an enhanced review of the literature on assessment practice was made available. The review of assessment literature, entitled Assessment Bibliography, includes citations concerning assessment practice and has been bolstered by a detailed bibliography of assessment-related research conducted at UTK. During the final year of the project, the Center for Assessment Research and Development (the new parent organization for the Assessment Resource Center) assumed the role of editing Assessment Update, the first newsletter devoted to outcomes assessment. The Director of the Center serves as executive editor of the newsletter and the Associate Director provides a quarterly column on assessment measures. This responsibility has provided the Center with additional opportunities to identify resource people and disseminate information about assessment.

For the future, work continues on a revision of the Bibliography of Assessment Instruments. The new bibliography will contain more information on institutional experiences using various assessment instruments. In addition, the Associate Director's column on assessment measures will provide another means of disseminating information about institutional experiences with assessment instruments.

The current interest in measures of general education outcomes, coupled with UTK's extensive experience with the ACT-COMP exam, has motivated the Center's staff to develop a summary of research on the COMP exam. In addition, the staff is preparing an updated bibliography of assessment resource people and institutions. Excerpts from the summary of research on the ACT-COMP exam are included in Appendix C.

Objective 2: Gain recognition throughout the academic community. In order to attain the second program objective, gaining recognition for the Assessment Resource Center, staff members relied on three methods of enhancing the Center's visibility: (1) making invited presentations at national and regional meetings; (2) presenting refereed papers at national and regional meetings; and (3) publishing articles on assessment. During the first year of the project much of the effort to gain recognition for the Center was directed toward making invited presentations. Forums for these presentations included the annual meetings of the American Association for Higher Eduction (AAHE), the American Educational Research Association (AERA), the Association for the Study of Higher Education (ASHE), the Association of American Colleges (AAC), the AAHE Assessment Forum, and ten other national organizations. Not only did these meetings offer an opportunity to announce the creation of the Assessment Resource Center, they also provided lists of participants that later were used in producing a mailing list for the brochure describing the ARC. During the first year, the project director continued to publish articles on assessment and involved the associate director in two of these articles.

During the second year of the project, the Center's staff continued to make invited presentations and present competitively selected papers at annual meetings of the AAHE, AERA, AAC, the American Evaluation Association (AEA), the AAHE Assessment Forum, the FIPSE project directors, and approximately twenty other organizations. Participation in the OERI assessment audit and the editorship of the New Directions volume provided additional opportunities to describe the activities of the Center in published materials. The description of the role of the Center in supporting the assessment audit provided by Clifford Adelman in the introduction to Performance and Judgment and references to the Center's operations in the New Directions volume have helped to establish the ARC as a leader in the assessment movement.

As in the second year of the Center's operation, the third year of the project included presentations at the annual meetings of the AAHE, AERA, AEA, the Association for Institutional Research (AIR), the FIPSE project directors, and at least twenty other organizations. In addition, the Center joined the electronic assessment bulletin board (Assessnet) and was listed by ERIC as a clearinghouse on assessment. Competitively selected papers and published articles also played an important role in enhancing the credibility of the Center.

For the future, the Center's staff is focusing on presenting competitively selected papers at national meetings (AAHE, AEA, AERA, AIR) as a means of maintaining the current leadership position of the ARC (now Center for Assessment Research and Development). For example, staff members have proposed 11 papers, symposia, and/or workshops for these four national meetings in 1990-91.

Objective 3: Provide information about assessment materials and procedures. Efforts to enhance the visibility of the Center also serve an educative function and help to meet the third program objective, the dissemination of information. Between 1986 and 1989, presentations by the Center's staff have changed from descriptions of the assessment program at UTK to discussions of various paradigms for assessing student outcomes.

In addition to making presentations, the ARC has responded to the need for information in two other ways: (1) establishing contact with interested parties by mail, telephone, and personal visits; and (2) conducting national workshops. During the three years of the project, the number of responses to requests for information and the number of workshop attendees has increased steadily. Over these three years two significant changes have occurred in the handling of direct contacts. First, from years one to three there has been a decrease in the number of contacts with visitors to the UTK campus and a much greater reliance on contacts by mail and telephone. The second change is that requests for reports and other assessment materials have increased. At the end of the second year of the project it became apparent that attempting to provide all of the materials requested would exhaust the Center's photocopying budget. As a result, individuals and institutions requesting information now are charged a fee sufficient to cover the costs of photocopying, mailing, and handling.

Attendance at the ARC workshops has grown steadily during the three years of the project. Approximately 120 people attended the first Strategies for Assessing Outcomes workshop, and by the final year of the project more than 200 people attended the workshop. These workshops provide an introduction to assessment for institutions that are considering establishing assessment programs or are in the initial stages of implementing outcomes assessment. The content of the workshops has emphasized applied topics and attempted to provide ample opportunities for interaction between workshop presenters and participants. Schedules for the three workshops are included in Appendix D.

Objective 4: Contribute to the discussion and solution of assessment issues and problems. In order to meet the fourth program objective and advance the practice of assessment, the ARC has organized specialized seminars/workshops on specific aspects of assessment each year. During the first year the Center organized what was expected to be a small workshop for institutional research staff on the institutional effectiveness criterion proposed by the Southern Association of Colleges and Schools (SACS). The workshop was intended to provide a showcase for solutions to assessment problems developed by the Offices of Institutional Research and Information Systems at UTK. This meeting, planned for 35-40, unexpectedly grew to a size of nearly 100 participants.

Based on the experience with an open-enrollment seminar during the first year of the project, subsequent seminars were limited to a small number of invited participants. The topic for the seminar conducted during the second year was institutional experience with the ACT-COMP exam. Representatives from eight institutions participated in a round-table discussion with members of the ACT-COMP staff. This discussion was helpful in communicating campus concerns about the COMP exam and served to emphasize the need for ACT staff to provide more technical information about the test.

During the third year of the project two seminars were held. The first dealt with procedures for developing institution-specific outcomes measures. Held in Princeton, New Jersey, this conference included representatives from seven institutions and drew on the expertise of the Educational Testing Service (ETS). One of the outgrowths of this expertise of the Educational Testing Service (ETS). One of the outgrowths of this meeting was a series of suggestions for developing tests in students' academic majors that subsequently was published by the ARC Associate Director in Assessment Update.

The second conference held during the final year of the project took place at Alverno College in Milwaukee in February 1989. At this conference Stephen Dunbar from the Lindquist Center for Educational Measurement at the University of Iowa reviewed experiences with assessment at the K-12 level. Dunbar's conclusion that a primary weakness of K-12 assessment is that data are not used for program improvement helped provide the impetus for UTK's new FIPSE project. This conference produced the lead article for the Fall issue of Assessment Update. Lists of participants at all of the ARC seminars are presented in Appendix E.

During the third year of the project the scope of activities for ARC staff was broadened to include an international dimension. A British consulting firm, H+E Associates, invited the University of Tennessee to co-sponsor an international conference on assessing quality in higher education. This event was held at Robinson College of Cambridge University in July 1989. In addition, FIPSE funds were used to support the work of John Harris, professor at David Lipscomb University in Nashville, TN, as the U.S. representative on the international Study Group on Evaluation of Higher Education sponsored by the Organization for Economic Cooperation and Development (OECD) in Paris. In April 1989 the project director contributed one of the background papers for the Study Group.

RESULTS

The Assessment Resource Center has effectively accomplished each of its four objectives.

Objective 1. In addition to the Bibliography of Assessment Instruments and the listing of articles on assessment practice (Assessment Bibliography), the Center has produced 8 additional major reference works and 19 research reports. Tables of contents for the major reference works and a listing of the research reports are included in Appendix F. Edited versions of the bibliographies have been published in Performance and Judgment and the New Directions in Institutional Research volume, Implementing Outcomes Assessment: Promise and Perils. Copies of the chapters from Performance and Judgment are included in Appendix G.

Audience response to these materials has been very positive. Through September 15, 1989, individuals representing 158 institutions had requested copies of the latest versions of these materials. A detailed description of the characteristics of these requests is presented in Tables 1-3 of Appendix H. A survey of users was conducted in Spring 1989 for the purpose of assessing perceived quality of the Center's publications. The scale used for the rating was a 5-point Likert Scale ranging from 1 = poor quality to 5 = excellent quality. The average rating for all publications was 4.10, and the range of ratings was from 3.48 for Material on Locally Developed Tests to 4.49 for the Assessment Bibliography. Results of the evaluations of ARC publications are provided in Table 4 of Appendix H.

Objective 2. Since 1986, the Center's staff have made 85 invited and/or competitively-selected presentations on assessment. The number of articles published on assessment has increased from 5 in 1987 to 7 in 1988 to 12 in 1989. Center staff have visited campuses in 24 states and Puerto Rico to make presentations on assessment and to assist faculty in developing their own assessment plans.

Objective 3. To date the ARC has responded to requests for information and distributed information to colleges and universities in 49 states, the District of Columbia, Puerto Rico, and five foreign countries. A recent review of the quality of the Center's efforts reveals that ARC users give a mean quality rating of 4.03 (on a 5-point scale) to these efforts. Results of the evaluations of the ARC's communication efforts overall are presented in Table 5 of Appendix H.

As previously noted, attendance at the Center's workshop has increased from 120 participants representing 20 states at the March 1987 workshop to 126 participants from 31 states at the November 1987 workshop to 204 participants from 26 states at the November 1988 workshop. Analyses of workshop evaluations indicate that participants have consistently rated the overall quality of the workshops between 7 and 8 on a 10-point scale. Table 6 in Appendix H presents the participants' ratings from each of the ARC workshops.

Objective 4. Evaluations of the quality of the seminars designed to improve assessment practice are more subjective. However, participants in these seminars have consistently indicated that the material presented was very useful. This can be seen in the evaluations of the workshop held in August 1987, where participants rated the overall quality of the workshop 7.3 on a 10-point scale. Ratings for this workshop are included in Table 7 of Appendix H. The letters of support from participants attending the three small workshops (see Appendix I) also provide evidence of the quality of these conferences.

The International Conference on Assessing Quality in Higher Education at Cambridge University drew 45 participants representing 9 countries. The event was judged highly successful by all who were involved, and H+E Associates invited the University to co-sponsor a second such conference in 1990.

John Harris completed his work with the OECD Study Group on Evaluation of Higher Education and submitted his final report to FIPSE in November 1989. An excerpt from that report appears in Appendix J.

PLANS FOR CONTINUATION AND DISSEMINATION

Most of the essential activities of the Assessment Resource Center are being continued under the auspices of the new UTK Center for Assessment Research and Development (CARD). The Bibliography of Assessment Instruments, the Assessment Bibliography, and the collection of assessment-related research reports will be continuously updated and disseminated. The first "Strategies for Assessing Outcomes" workshop supported solely by CARD was held in Knoxville on November 6-7, 1989. The 182 enthusiastic participants, among whom were panelists from Alverno College, Northeast Missouri State University, and Miami-Dade Community College, as well as the University of Tennessee, Knoxville, provided evidence of the continuing viability of a medium-sized, highly interactive workshop that will address the questions and concerns of the newest practitioners of campus-based outcomes assessment.

CARD staff have an abiding interest in continuous improvement of their activities, and will continue to pay close attention to careful evaluations of their materials and services. Continuing a practice begun at the first "Strategies" workshop in 1987, the November 1989 workshop in Knoxville was concluded with the administration to all participants of an evaluation form.

The demand for the printed materials developed by the Center continues to increase as the number of assessment practitioners grows and as more individuals and institutions learn of the Center's services.

A new FIPSE grant received by CARD in September 1989 will enable representatives of seven diverse institutions in Tennessee to jointly develop and administer a survey for employers of the graduates of these institutions. Simultaneously, staff at the seven institutions will begin to study the implications of W. Edwards Deming's quality improvement philosophy for higher education.

The one area of the Center's services that, unfortunately, cannot continue without external funding is the exploration of assessment issues by the consortium of campus-based assessment leaders established under ARC auspices. At its final meeting during the Assessment Forum in Atlanta in June 1989, this group outlined a series of some half-dozen topics for discussion that it wished to pursue. However, since campus funding for academic travel is limited, the group concluded that it could not meet again for an extended special session with a consultant unless another source of funding can be found.

The responsibility for editing Assessment Update should maintain the visibility of Center staff and assist us in continuing to provide leadership with respect to the future development of the field of assessment in higher education. Moreover, sponsorship of a series of international conferences -- the second to be held at St. Andrews University in Scotland in July 1990 -- will extend staff association with assessment to an international level.

SUMMARY AND CONCLUSIONS

When the concept of an Assessment Resource Center was proposed in 1986, there was no clear understanding of the extent to which assessment would continue to be an important national and international issue. We wondered if the introductory "Strategies" workshop would be needed after the first or second year of our work.

Now it is obvious that the public interest in accountability in postsecondary education is sufficiently great that assessment will be an issue with high priority for years to come. At the November 1989 "Strategies for Assessing Outcomes" workshop in Knoxville, there was every indication that participants would welcome a similar offering in 1990. One of our most important sources of participants each year has been the colleagues of participants in prior years.

The demand for materials and assistance in connection with the practice of campus-based outcomes assessment is another indication of the continuously increasing nature of interest in assessment. The very favorable reaction of colleagues around the world to the ARC materials has given us the impetus to continue to develop the collection. In addition, we hope that we will continue to be included in national and international discussions of the future of assessment and can make a significant contribution to the literature of assessment. The large number of students involved in assessment, the diversity of departments that have undertaken their own assessment activities, and the involvement of faculty in a variety of disciplines in the pursuit of assessment-related research virtually assure the continued leadership role of the University of Tennessee, Knoxville in the advancement of knowledge about the practice of outcomes assessment. We hope to use that role to chart a wise course for assessment research and development in future years.

APPENDIXES

APPENDIX A

Overview of the Assessment Resource Center

9/15/86 - 9/15/89

The Assessment Resource Center
at the
University of Tennessee, Knoxville

Funded by the University and a grant from
The Fund for the Improvement of
Postsecondary Education
U.S. Department of Education

Purpose: To provide leadership and assistance for colleges and universities throughout the country interested in assessing the outcomes of higher education.

A Summary of Activities of the Center Staff & University Colleagues

* Preparation of collections of useful materials, bibliographies of assessment instruments, and summaries of successful institutional practices

* Publication of numerous articles, several books, and a national newsletter (in association with Jossey-Bass)

* Distribution of information to 1400 colleges, universities, and organizations in 49 states and 6 other countries

* Conduct of workshops and seminars

-March 1987 workshop in Nashville - 120 participants from 20 states
-August 1987 workshop in Knoxville - 95 participants from 12 states
-November 1987 workshop in Memphis - 126 participants from 31 states
-November 1988 workshop in Knoxville - 204 participants from 29 states

* Presentations at international, national and regional meetings

-ACT Conference Series on Assessment
-American Association for Higher Education
-American Association of State Colleges and Universities
-American Educational Research Association
-Association for Handicapped Student Services Programs in Postsecondary Education
-Association for Institutional Research
-Association for the Study of Higher Education
-Association of American Colleges
-Bryn Mawr Summer Institute for Women in Higher Education Administration
-California State University System
-Eastern Sociological Association
-John Dewey Society
-National Association of Student Personnel Administrators
-National Association of Colleges and Teachers of Agriculture
-National Conference on Assessment in Higher Education
-National Education Association
-Organization for Economic Cooperation and Development - Paris
-Society for College and University Planning
-Southeastern Association for Institutional Research
-Southern Association for Institutional Research
-Southern Association of Colleges and Schools

* Consultation with national and regional organizations

-Administrators of Accounting Programs
-American Association for Higher Education
-American College Testing Program
-ASHE/ERIC - Consulting Editor
-Association for the Study of Higher Education
-Association of American Colleges
-Kansas City Regional Council for Higher Education
-National Association of State Universities and Land-Grant Colleges
-Pennsylvania Council of Academic Deans
-Southern Association of Colleges and Schools
-UCLA-FIPSE Value-Added Consortium
-U.S. Department of Ed. - FIPSE and Office of Ed. Research & Improvement

* Consultation with colleges and universities in 21 states, including:

-California State University, Los Angeles
-University of Alabama, Tuscaloosa
-University of Mississippi
-University of Missouri, Columbia
-University of North Carolina, Chapel Hill
-University of Wisconsin, Madison

* Conduct of campus visits for representatives of institutions in:

Australia Canada The Netherlands West Germany 14 States

* Conduct of special topics seminars for assessment leaders:

-April 1988 in Kansas City - "Users' problems in using the ACT COMP exam" for institutional representatives from 7 states and ACT COMP staff

-October 1988 at ETS in Princeton, N.J. - "Problems in developing tests in the major" for 11 assessment leaders

-February 1989 - "Implications of K-12 competency testing for assessment in higher education" for 12 assessment leaders

** Recognition Received by UTK Assessment Program

0 National Council on Measurement in Education Triennial Award for Excellence in Using Educational Measurement Technology (1984)

0 Selected to present the United States Case Study on the Assessment of Institutional Effectiveness at the annual meeting in Paris of the Organization for Economic Cooperation and Development (1986)

0 Proposal for Assessment Resource Center was one of 75 selected for funding by FIPSE from 2100 proposals submitted (1986)

0 Selected as sole source contractor for the Assessment Audit funded by the U.S. Department of Education (1987)

0 Director honored by the American Association for Higher Education for contributions to the field of assessment

0 Invited by Jossey-Bass Publishers of San Francisco to edit the nation's first quarterly publication in the field of assessment (1988)

0 Featured in CNN interview (1987)

0 Participant in national teleconference on assessment (1988)

0 Subject of features published in Le Monde, the New York Times, the Chronicle of Higher Education, Change, and the international Journal of Institutional Management in Higher Education

Quick Summary of Published Work on Assessment by ARC Personnel

1984

- 0 Article in Educational Measurement

1985

- 0 Chapter in New Directions for Higher Education volume
- 0 Chapter in New Directions for Institutional Research

1986

- 0 Book: Performance Funding in Higher Education: A Critical Analysis
- 0 Article in Educational Record
- 0 Article in Conference Proceedings of meeting co-sponsored by AASCU and George Mason University
- 0 Article in Journal of Institutional Management in Higher Education

1987

- 0 Article in The Chronicle of Higher Education
- 0 Chapter in New Directions for Higher Education
- 0 Article in State Education Leader
- 0 Article in Virginia Community College Journal
- 0 Article in Yearbook of American Universities

1988

- 0 Article in Journal of Higher Education
- 0 Edited volume in New Directions for Institutional Research series
- 0 Chapter in Performance and Judgment published by the U.S. Office of Ed. Research & Improvement
- 0 Article in Research in Higher Education
- 0 Article in Proceedings of the annual meeting of the John Dewey Society
- 0 Chapter in Yearbook of American Universities and Colleges. New York: Garland
- 0 Consultant for article in Changing Times
- 0 Editor for quarterly newsletter on assessment for Jossey-Bass Publishers, San Francisco

1989

- 0 Several articles in Assessment Update: Progress, Trends, and Practices in Higher Education
- 0 Chapter in The Theory and Practice of Outcomes Assessment. S. J. Reithlingshoefer (Ed.)
- 0 Two articles in Research in Higher Education
- 0 Article in Review of Higher Education Research
- 0 Proceedings of the 1989 International Conference on Assessing Quality in Higher Education

CAMPUS VISITS
1985-1989

Alabama
- Samford University - 1989

California
- Cal State, Los Angeles - 1987

Connecticut
- University of Connecticut - 1988

Georgia
- Berry College - 1988
- Shorter College - 1989

Hawaii
- University of Hawaii, Manoa - 1988
- Community College System - 1988

Illinois
- Rosary College - 1988
- Triton College - 1985

Indiana
- University of Southern Indiana - 1987

Kentucky
- Berea College - 1987
- Murray State University - 1988
- University of Louisville - 1986
- Kentucky State University - 1988

Louisiana
- Louisiana Tech - 1989

Michigan
- Saginaw Valley State College - 1985
- Wayne State University - 1986

Missouri
- Northeast Missouri State University - 1985
- Southwest Missouri State University - 1988

Mississippi
- Mississippi University for Women - 1987
- Jackson State University - 1988

New Jersey
- Kean College - 1987, 1988, 1989

North Carolina
- University of North Carolina, Charlotte - 1987
- Fayetteville State University - 1987
- Forsyth Technical Community College - 1989

Ohio
- University of Dayton - 1985
- University of Toledo - 1988

Oklahoma
- Northeastern State University - 1989

Pennsylvania
- Gannon University - 1985
- King's College - 1985
- Philadelphia College of Pharmacy & Science - 1989
- University of Scranton - 1989

Puerto Rico
- Humacao University College of the University of Puerto Rico - 1989

Rhode Island
- Rhode Island College - 1985, 1987, 1989
- Rhode Island Community College - 1989

South Carolina
- South Carolina State College - 1987
- Winthrop College - 1988
- University of South Carolina - 1989
- Midlands Technical College - 1989

South Dakota
- University of South Dakota, Vermillion - 1986

Tennessee
- Memphis State University - 1988
- Roane State Community College - 1987
- Milligan College - 1989
- University of TN, Chattanooga - 1989

Texas
- Southwest Texas State University - 1985
- University of Houston - 1987

Virginia
- Clinch Valley College - 1987, 1989
- James Madison University - 1985, 1989
- Old Dominion University - 1986
- Marymount University - 1988
- Virginia Highlands Community College - 1988
- Mountain Empire Community College - 1988
- Wytheville Community College - 1987

Washington
- Central Washington State University - 1987

APPENDIX B

Bibliography of Resource Persons and Institutions

Before beginning assessment programs, institutions should carefully examine the literature on assessment practice, including the experiences of other institutions.

An Annotated Bibliography and Program Descriptions

Gary R. Pike

Over the last five years, interest in assessing student educational outcomes has increased dramatically in this country. This growing interest has been fueled in large part by state laws and accreditation requirements. There has also been a parallel growth in the literature on assessment. In view of the variety of information currently available, this bibliography is not exhaustive. Instead, it provides a starting point for the study of assessment.

Sources on Assessment

Adelman, C. P. (ed.). *Assessment in American Higher Education: Issues and Contexts.* Washington, D.C.: U.S. Government Printing Office, 1985.

This series of essays, based on presentations made at the First National Conference on Assessment in Higher Education, examines assessment from a variety of perspectives, including the philosophy of assessment, assessment in professional/technical schools, selection of assessment instruments, and evaluating the costs of assessment.

Astin, A. W. *Achieving Educational Excellence: A Critical Assessment of Priorities and Practices in Higher Education.* San Francisco: Jossey-Bass, 1985.

T. W. Banta (ed.). *Implementing Outcomes Assessment: Promise and Perils.*
New Directions for Institutional Research, no. 59. San Francisco: Jossey-Bass, Fall 1988.

Astin examines the concept of quality in higher education. He begins by identifying some limitations of the traditional indicators of quality, and he advocates the use of a talent-development (value-added) approach to assessing quality. Using data from the Cooperative Institutional Research Program (CIRP), Astin suggests several steps institutions can take to promote educational quality.

Banta, T. W. (ed.). *Performance Funding in Higher Education: A Critical Analysis of Tennessee's Experience.* Boulder, Colo.: National Center for Higher Education Management Systems, 1986.

Tennessee's performance-funding initiative serves as the focal point of the essays in this volume. The topics include policy issues (such as the development of performance-funding criteria and the use of data in institutional planning) and measurement issues (such as using surveys to measure student satisfaction, and testing achievement in general education and in the major).

Bergquist, W. H., and Armstrong, J. R. *Planning Effectively for Educational Quality: An Outcomes-Based Approach for Colleges Committed to Excellence.* San Francisco: Jossey-Bass, 1986.

In providing a model for improving educational quality, the authors urge institutions to examine their missions, develop pilot programs designed to assist in accomplishing those missions, assess the effectiveness of the pilot programs, and then implement large-scale programs. They stress the importance of incorporating outcomes data into the planning process.

Berk, R. A. (ed.). *Performance Assessment: Methods and Applications.* Baltimore, Md.: Johns Hopkins University Press, 1986.

This volume is a basic reference work on performance assessment. Authors of the essays present a variety of methods, ranging from behavior rating to assessment centers. The authors also show uses of performance assessment in business, medicine, law, teaching, and evaluation of communications skills.

Boyer, C. M., Ewell, P. T., Finney, J. E., and Mingle, J. R. "Assessment and Outcomes Measurement: A View from the States." *AAHE Bulletin,* 1987, *39* (7), 8–12.

These authors report the results of a survey conducted by the Education Commission of the States. The purpose of the survey was to identify trends in state-promoted assessment. Results showed that approximately two-thirds of the states have established or are establishing programs that provide incentives for assessment efforts.

Educational Testing Service. *Assessing the Outcomes of Higher Education: Proceedings of the 1986 ETS Invitational Conference.* Princeton, N.J.: Educational Testing Service, 1987.

Topics addressed in the papers presented at this invitational conference on assessment include the responses of accrediting associations, state agencies, and colleges and universities to the assessment movement; the use of unobtrusive measures in assessing student outcomes; and the role of value-added analyses in assessing educational outcomes.

El-Khawas, E. "Colleges Reclaim the Assessment Initiative." *Educational Record,* 1987, *68* (2), 54–58.

In this article, the author summarizes the results of a recent survey conducted by the American Council on Education (ACE). The author argues that colleges and universities are taking the lead in promoting the assessment of student educational outcomes. According the ACE survey, colleges and universities are developing campus assessment programs designed to improve campus planning, rather than to satisfy external mandates.

Ewell, P. T. "Assessment: Where Are We?" *Change,* 1987, *19* (1), 23–28.

While Ewell briefly examines efforts of private institutions to assess student outcomes, the primary focus is on the response of public institutions to state mandates. Ewell describes several state-mandated approaches and institutional responses to them, as well as several concerns arising from the trend toward state-mandated assessment.

Gronlund, N. E. *Constructing Achievement Tests.* (3d ed.) Englewood Cliffs, N.J.: Prentice-Hall, 1982.

The author provides a basic introduction to constructing achievement tests. Gronlund discusses all phases of test construction and evaluation, including specification of educational objectives, development of test items, and evaluation of test items. Gronlund also identifies differences in construction and scoring between objective and essay examinations.

Henerson, M. E., Morris, L. L., and Fitz-Gibbon, C. T. *How to Measure Attitudes.* Newbury, Park, Calif.: Sage, 1983.

These authors describe a variety of approaches for measuring attitudes, including surveys, interviews, and observations of behavior. They also discuss development, reliability, and validity of instruments.

Marchese, T. J. "Third Down, Ten Years to Go." *AAHE Bulletin,* 1987, *40* (4), 3–8.

Marchese traces the assessment movement over the last three years. The author describes six approaches to assessing educational outcomes:

the assessment center, assessment as learning, assessment as program monitoring, assessment of student development, assessment as standardized testing, and assessing through a senior examiner.

Representative Assessment Programs

The growing interest in educational outcomes has produced a variety of approaches to assessment. The following organizations engage in ongoing assessment. This list represents only a sampling of current programs, but it covers the variety of approaches being used at this time.

Alverno College
Judeen Schute, Alverno College, 3401 South 39th Street, Milwaukee, WI 53215; (414)382-6000. Designs and implements general educational outcomes assessment.

American Association for Higher Education
Patricia Hutchings, American Association for Higher Education, One Dupont Circle NW, Suite 600, Washington, DC 20036; (202)293-6440. Convenes annual forum, supports descriptive studies of assessment, and provides referral service.

Association of American Colleges
Carol Schneider, Association of American Colleges, 1818 R Street NW, Washington, DC 20009; (202)387-3860. Coordinates assessment programs that rely on visiting examiners.

City University of New York, Research Foundation
Harvey S. Wiener, Professor of English, CUNY/Research Foundation, 309 Clearview Lane, Massapequa, NY 11758; (516)799-1951. Assessment of word processing and writing effectiveness.

Educational Testing Service
Roy Hardy, Director, Educational Testing Service, 250 Piedmont Avenue NE, Suite 1240, Atlanta, GA 30308; (404)524-4501. Develops item banks for assessment of learning outcomes in five disciplines.

Harvard Medical School
Gordon Moore, Director, New Pathway Project, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115; (617)732-0634. Conducts comparative assessments of traditional and medical school curriculum models.

Harvard University
Richard Light, Professor, School of Education, Harvard University, Cambridge, MA 02138; (617)495-1183. Manages cooperative pilot project involving assessment at selective institutions.

Indiana University of Pennsylvania

Robert Millward, Director, Pre-Teacher Assessment Center, Indiana University of Pennsylvania, 136 Stouffer Hall, Indiana, PA 15705; (412)357-2480. Establishes pre-teacher assessment in new center that evaluates teaching abilities using classroom simulations.

James Madison University

Dary Erwin, Office of Student Assessment, James Madison University, Harrisonburg, VA 22801; (703)568-6211. Conducts assessment in seven broad areas: major, general education, interdisciplinary objectives, affective development, functional skills, alumni, and environment.

Kean College

Michael Knight, Donald Lumsden, Assessment of Student Learning and Development, Kean College of New Jersey, Union, NJ 07083; (201)527-2000. Uses faculty-developed outcomes assessment in each program area.

King's College

D. W. Farmer, Vice-President and Dean of Academic Affairs, King's College, Wilkes-Barre, PA 18711; (717)826-5900. Uses outcomes-oriented curriculum, complemented by course-embedded assessment program; emphasis on "transferable skills of liberal learning" linked with progress in major.

Miami University of Ohio

Karl Schilling, Associate Dean, Western College Program, Miami University, Oxford, OH 45056; (513)529-1809. Employs comparative assessment of discipline-based and interdisciplinary undergraduate curricula at Miami University of Ohio.

Northeast Missouri State University

Charles J. McClain, President; Darrell Krueger, Dean of Instruction; Administration/Humanities Building, Northeast Missouri State University, Kirskville, MO 63501; (816)785-4100. Employs value-added approach using standardized tests and uses surveys to assess student growth and evaluate the university.

Ohio Board of Regents

Elaine H. Hairston, Vice-Chancellor, Academic and Special Programs, Ohio Board of Regents, 30E Broad Street, 36th Floor, Columbus, OH 43266; (614)466-6000. Promotes excellence, stimulates assessment, and communicates program and institutional improvements to external agencies, on the basis of a statewide assessment project.

Rhode Island College

William Enteman, Provost and Vice-President for Academic Affairs, Rhode Island College, Providence, RI 02908; (401)456-8003. Develops assessment activities linked to curriculum revisions, new advising systems, and individualized educational plans.

South Dakota State University

Kris Smith, Assessment and Testing Office, South Dakota State University, Administration Building, Room 215, Brookings, SD 57007; (605)688-4217. Reviews general educational outcome assessment instruments after trial use; also uses a variety of subject area tests and surveys.

Southern Association of Colleges and Schools

Carol A. Luthman, Assistant Executive Director, Commission on Colleges, Southern Association of Colleges and Schools, 795 Peachtree Street NE, Atlanta, GA 30365; (404)847-6120. Develops manuals describing college use of outcomes assessment during the accreditation process.

State University of New York, Plattsburg

Thomas Moran, Assistant Vice-President for Academic Affairs, SUNY/Plattsburg, Plattsburg, NY 12901; (518)564-2080. Develops new assessment procedures as alternatives to nationally standardized tests.

Texas College and University System

Mary Griffith, Project Director, Community College and Technical Institute Division, Coordinating Board of the Texas College and University System, P.O. Box 12788, Houston, TX 78711; (512)475-0718. Defines college-level skills in reading, writing, and mathematics for subsequent adoption by Texas postsecondary system.

University of Kentucky

Charles Elton, Karen Carey, College of Education, University of Kentucky, 111 Dickey Hall, Lexington, KY 40506; (606)257-2627. Studies value-added approaches to assessing institutional effectiveness.

University of Massachusetts

José P. Mestre, Department of Physics and Astronomy, University of Massachusetts, Amherst, MA 01003; (413)545-2040. Researches and develops computer-assisted problem-solving skills (bilingual: English and Spanish).

University of New Mexico

Scott Obenshain, School of Medicine, University of New Mexico, P.O. Box 508, Albuquerque, NM 87131; (505)277-4823. Currently developing a self-assessment center for medical students.

University of Tennessee, Knoxville

Trudy W. Banta, Gary R. Pike, The Assessment Resource Center, University of Tennessee, 2046 Terrace Avenue, Knoxville, TN 37996-3504; (615)974-0883. Conducts a comprehensive assessment program involving testing in the major and in general education, and surveying students and alumni. Results are used in program reviews and institutional planning and improvement. Conducts workshops and disseminates information about assessment to other colleges and universities.

Gary R. Pike is associate director of the Assessment Resource Center at the University of Tennessee, Knoxville.

APPENDIX C

Research on the College Outcome Measures Project

INTRODUCTION

In 1976, the American College Testing Program (ACT) organized the College Outcome Measures Project (COMP) to develop a measure of "knowledge and skills relevant to successful functioning in adult society" (Forrest, 1982, p. 11). Available since 1979-80, the COMP exam has been administered at least once at more than 500 colleges, and it is used annually by approximately 100 four-year institutions in the evaluation of their general education programs (American College Testing Program, 1987). Until recently, the COMP exam was the only instrument designed for evaluating general education programs, and in 1989 it remains the only measure for which a substantial amount of data is available.

The COMP exam is available in two forms: the Objective Test (consisting of multiple-choice items) and the Composite Examination (containing multiple-choice questions, along with exercises requiring students to write essays and record speeches). ACT reports that the correlation between the two forms of the exam is .80, allowing the Objective Test to serve as a proxy for the Composite Examination (Forrest & Steele, 1982). Most institutions, including the University of Tennessee, Knoxville, use the Objective Test for program evaluation because it is easier to administer and score (Banta, Lambert, Pike, Schmidhammer, & Schneider, 1987).

The Objective Test takes approximately 2½ hours to administer and contains 60 questions, each with two correct answers. The questions are divided among 15 separately timed activities drawing on material (stimuli) from television programs, radio broadcasts, and print media. Students taking the COMP Objective Test are instructed that there is a penalty for guessing (i.e., incorrect answers will be subtracted from their scores), but that leaving a question blank will not be counted against them. The combination of two correct answers for each question, the guessing penalty, and no penalty for not answering a question means that the score range for each of the 60 items is from -2 to 2 points. A score of -2 represents two incorrect answers, while a score of -1 represents one incorrect answer and one left blank. A score of 0 can represent either both answers left blank or one correct and one incorrect answer. A score of 1 represents one correct answer and a blank, and a score of 2 represents two correct answers. For interpretability, these scores are rescaled (0 to 4), making the maximum possible score on the Objective test 240 points and a chance score 120 points.

New forms of the COMP Objective Test are developed on an annual basis. In order to ensure comparability of scores across forms, the COMP staff equates each new form to the original test (Form III). This equating is done using samples of high school and college seniors who are double-tested using Form III and the new form of the test. Statistical procedures involve the use of Angoff's (1984) Design II (Steele, J. M., personal communication, 14 September 1989).

In addition to a total score, the COMP Objective Test provides three content subscores (Functioning within Social Institutions, Using Science and Technology, and Using the Arts) and three process subscales (Communicating, Solving Problems, and Clarifying Values). In the technical manual for the COMP exam, ACT staff report that the alpha reliability of the total score is .84, and that reliability estimates for the subscores range from .63 to .68 (Forrest & Steele, 1982). Estimates of parallel-forms reliability are .79 for the total score and range from .53 to .68 for the subscores. More recently, Steele (1988) has argued that the reliability coefficients for group means on the COMP exam are .98 for the total score and range from .97 to .98 for the subscores.

Many colleges and universities are drawn to the COMP exam because it offers to provide objective evidence of student intellectual growth (value added) over the course of a college education. Students who persist at an institution can be tested upon entrance and again at the end of two or four years of college in order to determine the growth attributable to their educational experiences (Forrest 1982). Partly because many institutions are unwilling to wait two or four years to evaluate student learning, COMP staff provide an estimate of student gain. Based on the fact that the correlation between total scores on the Objective Test and entering ACT Assessment Composite scores is .70, the COMP staff have constructed a concordance table from which institutions may estimate mean freshman COMP scores if they have mean ACT Assessment Composite scores (or mean SAT scores) (Banta, et al., 1987). By subtracting the estimated freshman score from the actual score for graduating students, an estimate of score gain, or value-added, can be obtained.

The University of Tennessee, Knoxville (UTK) has the most extensive institutional COMP database in the country. Since 1980 several hundred seniors at UTK have been tested annually using the COMP Objective Test. In 1985, the test became a graduation requirement for every senior, and the test is annually administered to a sample of approximately one-half of the freshman class. When students take the COMP exam at UTK they are asked to complete an extensive survey designed to gather information about their previous educational experiences. These responses are combined with background data from student records and test scores to produce the UTK database. As of 1989, the UTK database on the COMP exam contains test scores, background data, and survey responses for more than 20,000 students.

Because UTK is in a unique position to provide evidence of the technical quality of the COMP exam, it has undertaken a variety of research projects. In conducting research on the COMP exam, it was impossible for UTK to divorce itself from the context in which COMP scores are used -- the Tennessee performance funding program. The results of a representative sample of these projects are included in this publication. As a guide, an annotated bibliography of this research is provided.

References

American College Testing Program. (1987). College Outcome Measures Program: 1987-88. Iowa City, IA: Author.

Angoff, W. H. (1984). Scales, norms and equivalent scores. Princeton, NJ: Educational Testing Service. (Reprint of a chapter in R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D. C.: American Council on Education, 1971).

Banta, T. W., Lambert, E. W., Pike, G. R., Schmidhammer, J. L., & Schneider, J. A. (1987). Estimated student score gain on the ACT COMP exam: Valid tool for institutional assessment? Research in Higher Education, 27, 195-217.

Forrest, A. (1982). Increasing student competence and persistence: The best case for general education. Iowa City, IA: ACT National Center for the Advancement of Educational Practice.

Forrest, A., & Steele, J. M. (1982). Defining and measuring general education knowledge and skills. Iowa City, IA: American College Testing Program.

Steele, J. M. (1988, May). Using measures of student outcomes and growth to improve college programs. Paper presented at the annual forum of the Association for Institutional Research, Phoenix, AZ.

ANNOTATED BIBLIOGRAPHY

Banta, T. W., Lambert, E. W., Pike, G. R., Schmidhammer, J. L., & Schneider, J. A. (1987). Estimated student score gain on the ACT COMP exam: Valid tool for institutional research? Research in Higher Education, 27, 195-217.

In this article, the authors examine issues related to the reliability and validity of a measure of the value added by education based on estimated score gains on the ACT COMP exam. Using data from the 1985 (N=843) and 1986 (N=2226) academic years, several problems with the reliability and validity of estimated gain scores are identified. Concerns related to reliability include: (1) the large standard deviation of gain scores; (2) the significant differences between estimated and actual gain scores; and (3) the unreliability of difference scores as indicators of change. Concerns related to validity include: (1) the significant negative correlation between pretest scores (ACT Assessment Composite) and estimated gain; (2) the systematic exclusion from estimates of gain of subgroups due to the absence of ACT Assessment scores for those groups; and (3) the presence of counter-intuitive relationships between estimated gain scores and measures of students' educational experiences.

Banta, T. W., & Pike, G. R. (in press). Methods of comparing outcomes assessment instruments. Research in Higher Education.

The purpose of this article is to outline a strategy for use by faculty in comparing the relative efficacy of outcomes assessment instruments in gauging program effectiveness. The methods described by the authors include: (1) asking faculty to compare the content of the instruments with objectives for the general education program; (2) asking students about their perceptions of the instruments; and (3) analyzing the psychometric properties of the instruments. Based on analyses of the ACT-COMP exam and the Academic Profile, the authors conclude that neither test covers more than 30 percent of the general education goals defined as important at UTK. Furthermore, students tend to rate both tests as poor measures of their general education experiences, and analyses of the psychometric properties of these tests indicate that both are primarily measures of individual differences.

Phillippi, R. H. (1989). A comparison of reliability and difficulty levels for three forms of the COMP exam (Research Report No. 89-02). Knoxville, TN: University of Tennessee, Center for Assessment Research and Development.

Based on data from Forms 7, 8, and 9 of the COMP Objective Test, the author concludes that the most likely response on a given item on the COMP exam by chance alone is one right and one wrong. A score on the test, therefore, is actually a measure of the number of items for which both responses are correct. Based on this type of scoring, a reliability analysis of Forms 7, 8, and 9 shows little change across forms. Analysis of item difficulties shows that all three forms of the test are excessively easy for seniors, although Forms 7 and 8 are reasonably difficult for freshmen. Form 8 of the test was the most difficult form of the test for seniors. However, seniors also scored highest on form 8. This paradox can be explained by the equating process used for the COMP exam. (UTK students score significantly higher than the equating population on this form of the test.) The net effect is a two-point gain in equated scores for students at UTK.

Phillippi, R. H. (1989). Analysis of the impact of social sciences coursework on ACT-COMP scores (Research Report No. 89-01). Knoxville, TN: University of Tennessee, Center for Assessment Research and Development.

The research presented in this report was designed to assess the impact of differential patterns of social science coursework on total score and all subscale scores of the ACT-COMP exam. Using COMP scores for more than 10,000 seniors tested at UTK, the author found that social science coursework significantly affects scores on the Functioning within Social Institutions, Using Science and Technology, Using the Arts, Communicating, and Clarifying Values subscales. Social science coursework did not influence Solving Problems scores. In every case, the magnitude of these effects is quite small. The author concludes: "It appears that the validity of total score on the ACT-COMP for measuring the effects of education is questionable. Education, as measured in coursework, seems to effect subscores on the ACT-COMP, but those effects seem to be counterbalanced by opposite effects with other subscores. All of this results in total scores which do not reflect differences on the basis of education."

Pike, G. R. (1989). A comparison of the College Outcome Measures Program (COMP) and the Collegiate Assessment of Academic Proficiency (CAAP) exams (Research Report No. 89-10). Knoxville, TN: University of Tennessee, Center for Assessment Research and Development.

Using the concepts of the substantive, structural, and external components of construct validity, the author investigates the use of the COMP and CAAP exams at UTK. Although the COMP exam was slightly superior to the CAAP in its coverage of general education content, neither test covered more than 30% of UTK's general education goals. In addition, neither test provides highly dependable measures of educational outcomes. This research also raises serious questions about the appropriateness of the scoring model used by these exams, both at the item level and in the construction of subscales. The most disappointing result of this research is the insensitivity of the COMP and CAAP exams to patterns of general education coursework. The variables that do predict successful performance on the COMP and CAAP exams are the students' levels of ability when they enter college, self-reports of how hard they try on the tests, and to a lesser extent, their gender and grade point averages in college.

Pike, G. R. (1989). Students' background characteristics, educational experiences, educational outcomes: A model for evaluating assessment instruments. In S. J. Reithlingshoefer (Ed.), Developing assessment partnerships between community and senior colleges: The theory and practice of outcomes assessment (pp. 79-91). Fairfax, VA: George Mason University.

The research presented in this paper investigates the criterion-related validity (instructional sensitivity) of the ACT-COMP exam. Two requirements for criterion-related validity are proposed and tested: (1) test scores should be more strongly related to educational experiences than to background characteristics; and (2) subscores should be differentially related to students' educational experiences. Two conclusions regarding the validity of the COMP exam emerge from this research on 2528 seniors: (1) while COMP scores are significantly related to the total amount of coursework, the explanatory power of this relationship, relative to the power of relationships between COMP scores and background characteristics, is quite small; and (2) subscores on the COMP exam are not differentially related to educational experiences, indicating that these subscores provide no unique information for program planning and evaluation.

Pike, G. R. (1989). The performance of black and white students on the ACT-COMP exam: An analysis of differential item functioning using Samejima's graded model (Research Report No. 89-11). Knoxville, TN: University of Tennessee, Center for Assessment Research and Development.

In this research report, the author describes the results of a study designed to evaluate differential item functioning for 471 blacks and 9237 white students who completed Form 8 of the COMP Objective Test. Using item response theory, the author found that 58% of the questions on the Objective Test produced significant levels of difficulty-shift in favor of whites. No instances of significant difficulty-shift in favor of blacks was found. The author notes that rates of differential item functioning are not evenly distributed across COMP subscales, and when items are categorized on the basis of content skills (identification versus explanation), questions designed to measure explanation produce higher rates of difficulty shift. In addition, the author suggests that questions which are based on unfamiliar stimuli may produce high rates of differential item functioning.

Pike, G. R. (1989). Background characteristics, college experiences, and the ACT-COMP exam: Using construct validity to evaluate assessment instruments. Review of Higher Education, 13, 91-117.

This article presents one criterion for judging the appropriateness of standardized tests as assessment instruments. To be valid indicators of program effectiveness, test scores should be related to factors associated with educational quality and should discriminate between these indicators and variables beyond the control of higher education (e.g., demographic characteristics). Research on the appropriateness of the ACT-COMP exam for program evaluation was conducted using data from 5936 seniors. This research provides mixed results. Specifically, results indicate that the ACT-COMP exam is a better measure of individual differences (entering academic ability/achievement) than of program quality. While the COMP exam is primarily a test of individual differences, it is sensitive to indicators of program quality (patterns of coursework) once effects for background characteristics are removed statistically.

Pike, G. R., & Banta, T. W. (1989, March). Using content validity to evaluate assessment instruments: A comparison of the ACT-COMP exam and the ETS Academic Profile. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Using Messick's work on construct validity, the authors examine the substantive, structural, and external components of test use by the Tennessee Higher Education Commission (THEC) and the University of Tennessee, Knoxville (UTK). Results indicate that both the COMP exam and the Academic Profile are equivalent in their coverage of UTK's general education goals. Regarding the structural component of test use, the Academic Profile is superior to the COMP exam in its ability to accurately differentiate among students/programs. Analysis of the structural component also reveals that both tests measure a single underlying construct, and analysis of the external component suggests that this construct is academic ability, not program quality. Analysis of data from the two tests also suggests that the COMP exam is somewhat more sensitive to educational effects than is the Academic Profile, but only after the effects of entering ability/achievement are removed.

Pike, G. R., & Phillippi, R. H. (1989). Generalizability of the differential coursework methodology: Relationships between self-reported coursework and performance on the ACT-COMP exam. Research in Higher Education, 30, 245-260.

The Differential Coursework Patterns Project (DCPP) appears to offer a method of linking outcomes measures to program data. However, questions about the generalizability of this method have been raised. This study indicates that the DCPP methodology can be used with at least two different tests, and that coursework measures can be gathered either through transcript analysis or students' self-reports. Specific findings of this research on 2943 seniors indicate that residuals for the six COMP subscores (produced by regressing subscores on ACT Assessment scores, age, and motivation scores) are related to coursework clusters. For example, coursework in the humanities tends to be related to residuals for Solving Problems, Clarifying Values, and Using the Arts, while calculus and physical science coursework tends to be related to residuals for Communicating and Using Science and Technology.

Pike, G. R., & Phillippi, R. H. (1989, May). Using generalizability theory in institutional research. Paper presented at the Forum of the Association for Institutional Research, Baltimore.

For assessment practitioners, errors of measurement are of critical concern because they can lead to incorrect judgments about the quality of education programs and suggest inappropriate strategies for program improvement. This paper describes the use of generalizability theory in identifying errors of measurement and indicates that variance in students and items should be considered sources of error when group means are the unit of analysis. Using a synthetic data set of 9000 respondents that is very similar to the national data compiled by ACT, the authors conclude that the use of group means (instead of individuals) as the unit of analysis will likely decrease the dependability of measurement, not increase it.

APPENDIX D

Schedules for Assessment Resource Center Workshops

TENTATIVE OUTLINE FOR NASHVILLE WORKSHOP: "Strategies for Assessing Outcomes"

Sponsored by the Assessment Resource Center
1987

Sunday, March 22

7:00 - 9:00 p.m.	REGISTRATION for early arrivals

Monday, March 23

8:00 - 8:45 a.m.	REGISTRATION and Coffee
8:45 - 9:00	WELCOME and Introductions--by Peter Ewell
	OVERVIEW OF ASSESSMENT
9:00 - 9:15	At Alverno College--by Georgine Loacker
9:15 - 9:30	At Northeast Missouri State University-- by Stuart Vorkink
9:30 - 9:45	At The University of Tennessee, Knoxville-- by Trudy Banta and Homer Fisher
9:45 - 10:30	INITIATING AN ASSESSMENT PROGRAM - Essential Elements Consultant Panel moderated by Peter Ewell
10:30 - 11:00	Questions from Participants about Initiating a Program
11:00 - 11:15	BREAK (Tables of materials provided by Center and by participants available for review)
11:15 - 12:30 p.m	ASSESSMENT METHODS - Concurrent Sessions
	Assessing Student Abilities - Georgine Loacker
	Assessing General Education Outcomes - Aubrey Forrest of ACT
	New Approaches from ETS - Roy Hardy of ETS
	Assessing Student Learning in the Major - Stuart Vorkink and Trudy Banta
	Using Survey Data in Institutional Planning - Homer Fisher and Bill Lyons of UTK
	Using Available Campus Data - Peter Ewell
	Filmed Interview with UTK Faculty Concerning Assessment - Gary Pike of UTK
12:30 - 1:45	LUNCHEON and Opportunities for Groupings by Institutional Type
2:00 - 3:15	CONCURRENT SESSIONS (Reprise)
3:15 - 3:30	BREAK
3:30 - 4:15	PANEL OF CONSULTANTS Summarizing Concerns Expressed in Concurrent Sessions
4:15 - 5:00	MEETINGS OF INSTITUTIONAL TEAMS to Discuss Plans for Assessment and Contribute Questions for Tuesday Morning Session (all consultants available for conferences).
5:00 - 6:15	RECEPTION (wine-cheese-fruit) - Meetings by Institutional Type
6:30	Free for Supper in Nashville

Tuesday, March 24

8:00 - 8:30 a.m.	CONTINENTAL BREAKFAST
8:30 - 10:00	CONSULTANT PANEL Addressing Various Topics (Some submitted by participants Monday evening) External incentives and internal benefits of assessment Motivating faculty and students to participate Organizing for data collection Converting data to information for planning Financing assessment activities The current politics of assessment
10:00 - 11:00	Consultants Respond to Additional Questions from Participants
11:00 - 11:15	BREAK
11:15 - 12:00	CONCURRENT SESSIONS "What Works in Comprehensive Universities" "What Works in Liberal Arts Institutions" (Previously identified participants will present brief reports on successful practices at their own institutions)
12:00 - 1:30	LUNCH
1:30 - 3:00	Opportunities for teams to confer with consultants (Optional activity)

USING OUTCOMES INFORMATION IN THE ASSESSMENT OF INSTITUTIONAL EFFECTIVENESS

-- WORKSHOP PROGRAM --

Thursday, August 27, 1987

11:00AM - 1:00PM	WORKSHOP REGISTRATION	Meeting Area Lobby
1:00 - 2:00	USING OUTCOMES INFORMATION TO ASSESS THE ACCOMPLISHMENT OF INSTITUTIONAL MISSION AND GOALS - Homer S. Fisher, Executive Vice-Chancellor for Business, Planning and Finance -&- Trudy W. Banta, Research Professor and Director, Assessment Resource Center	Cherokee "C"
2:00 - 2:30	SYSTEM ENVIRONMENT SUCCESS FACTORS - F. Ramsey Valentine, Director, Office of Information Systems	Cherokee "C"
2:30 - 3:15	USING EXISTING CAMPUS INFORMATION - John T. Hemmeter, Director, Office of Institutional Research	Cherokee "C"
3:15 - 3:30	REFRESHMENT BREAK	Meeting Area Lobby
3:30 - 4:15	GATHERING INFORMATION THROUGH SURVEYS - William Lyons - Professor, Political Science and Associate Director, Office of Institutional Research	Cherokee "C"
4:15 - 5:00	GATHERING AND USING TEST DATA: AN AUTOMATED APPROACH - Edward L. Medford, Senior Programmer/Analyst, Office of Information Systems	Cherokee "C"
5:00 - 5:15	WORK OF THE ASSESSMENT RESOURCE CENTER - Gary R. Pike, Assistant Director, Assessment Resource Center	Cherokee "C"
5:15 - 5:45	DISCUSSION	Cherokee "C"
6:00 - 7:00	RECEPTION FOR PARTICIPANTS, UTK FACULTY, AND ADMINISTRATORS	Great Smoky Mountain Center

Friday, August 28, 1987

8:00AM	- 8:30AM	CONTINENTAL BREAKFAST	Meeting Area Lobby
8:30	- 9:15	TECHNICAL CONSIDERATIONS IN USING SURVEYS AND EXISTING CAMPUS DATA - John T. Hemmeter, William Lyons, & Don S. Scroggins, Project Coordinator, Office of Institutional Research	Cherokee "C"
9:15	- 10:00	TECHNICAL CONSIDERATIONS IN USING TEST DATA - Edward L. Medford	Cherokee "C"
10:00	- 10:30	DISCUSSION	Cherokee "C"
10:30	- 10:45	REFRESHMENT BREAK	Meeting Area Lobby
10:45	- 11:15	MOTIVATING PARTICIPANTS AND REPORTING ASSESSMENT RESULTS - Trudy W. Banta	Cherokee "C"
11:15	- 11:45	RESEARCH USING ASSESSMENT DATA - Gary R. Pike,	Cherokee "C"
11:45	- 12:30PM	DISCUSSION	Cherokee "C"
12:30	- 2:00	LUNCHEON	Cherokee "A&B"

STRATEGIES FOR ASSESSING OUTCOMES
-- WORKSHOP SCHEDULE --

Sunday, November 8, 1987

7:00pm - 9:00pm	Early Registration (Hyatt Lobby)

Monday, November 9, 1987

8:00am - 8:30am	Registration
8:30 - 8:45	Welcome & Introduction - Peter Ewell
8:45 - 9:15	Brief Overview, Assessment at Four Institutions
9:15 - 10:15	Initiating an Assessment Program Panel Discussion Moderated by Peter Ewell
10:15 - 10:30	Refreshment Break
10:30 - 11:45	Assessment Strategies (Concurrent Sessions)
Noon - 1:15pm	Luncheon
1:30 - 2:45	Concurrent Sessions (Reprise)
2:45 - 3:00	Break
3:00 - 3:45	Summary of Questions/Concerns Expressed in the Concurrent Sessions
3:45 - 4:45	Meetings of Institutional Teams (Consultants Available)
5:00 - 6:15	Reception
6:30 -	Free for Supper in Memphis

Tuesday, November 10, 1987

8:00am - 8:30am	Continental Breakfast
8:30 - 10:00	Panel Discussion External Incentives and Internal Benefits Motivating Faculty and Students Organizing for Data Collection (Additional Topics may be Submitted by Participants)
10:00 - 10:15	Break
10:15 - 11:15	Concurrent Sessions (by Institutional Type/Size)
11:15 - Noon	Summary of Topics Discussed in the Concurrent Sessions and Workshop Summary by the Panel
Noon - 1:15	Luncheon

STRATEGIES FOR ASSESSING OUTCOMES
-- WORKSHOP SCHEDULE --

Sunday, November 13, 1988

7:00pm - 9:00pm	Early Registration (Holiday Inn Lobby)

Monday, November 14, 1988

8:00 - 8:30am	Registration
8:30 - 8:45	Welcome & Introduction
8:45 - 9:15	Brief Overview, Assessment at Four Institutions
9:15 - 10:15	Initiating an Assessment Program - Panel Discussion
10:15 - 10:30	Refreshment Break
10:30 - 11:45	Assessment Strategies (Concurrent Sessions) - Assessment as Learning - Does Assessment Make a Difference? A Case for Computer Adaptive Testing - Student Competences and Assessment in the General Education Core - The Campus Political Context for Assessment: Faculty, Staff, and Students - Organizing the Campus for Assessment - Using Surveys in Assessment - Using Available Campus Data - Developing Exams in the Major Field - Faculty Perspective
Noon - 1:15	Luncheon
1:30 - 2:45	Concurrent Sessions (Reprise)
2:45 - 3:00	Break
3:00 - 3:45	Summary of Questions/Concerns Expressed in the Concurrent Sessions - Panel Discussion
4:00 - 5:00	High Tea
5:00 - 6:00	Meetings of Institutional Teams and/or Regional Groups of Institutions (Panel Members Available)
6:00 -	Free for supper in Knoxville

Tuesday, November 15, 1988

8:00am - 8:30am	Continental Breakfast
8:30 - 10:00	Panel Discussion - External Incentives and Internal Benefits of Assessment - Motivating Faculty and Students to Participate - Organizing for Data Collection - Additional Topics to be Submitted by Participants
10:00 - 10:15	Break
10:15 - 11:15	Concurrent Discussion Sessions (by Institutional Type/Size) (Panel Members Available)
11:15 - Noon	Summary of Topics Discussed in the Concurrent Sessions and Workshop Summary by the Panel
Noon - 1:15	Luncheon

APPENDIX E

Participants at Three Small Seminars

April 1988 Kansas City Seminar Participants

Michael Field	Professor of English, Bemidji State University (MN)
Arnold Gelfman	Director of Institutional Research, Brookdale Comm. College (NJ)
Gary Moden	Office of Institutional Research, Ohio University
Linda Rudolph	Associate Vice President, Austin Peay State University (TN)
Howard Benoist	Vice President, Our Lady of the Lake College (TX)
Scott Olsen	Director of Institutional Research, Northeast Missouri State University
John Finney	Associate Dean, University of Puget Sound (WA)
Aubrey Forrest	ACT
Joe Steele	ACT
David Lutz	ACT
Trudy Banta	Director, Assessment Resource Center, UTK
Gary Pike	Associate Director, Assessment Resource Center UTK

October 1988 Princeton Seminar Participants

John Harris	Professor, David Lipscomb University
Tom Moran	Associate Vice President, SUNY Plattsburgh
Marcia Mentkowski	Director, Office of Assessment, Alverno College
Glen Rogers	Assistant Director, Office of Assessment, Alverno College
Michael Knight	Co-Director, Office of Assessment, Kean College (NJ)
Donald Lumsden	Co-Director, Office of Assessment, Kean College (NJ)
Tony Golden	Professor, Austin Peay State University (TN)
Dary Erwin	Director, Office of Assessment, James Madison University (VA)
Peter Ewell	Senior Associate, National Center for Higher Education Management Systems
Richard Burns	Vice President, ETS
Trudy Banta	Director, Assessment Resource Center, UTK
Gary Pike	Associate Director, Assessment Resource Center, UTK

February 1989 Milwaukee Seminar Participants

Don Farmer	Vice President, King's College (PA)
Peter Ewell	Senior Associate, NCHEMS
Michael Knight	Co-Director, Office of Assessment, Kean College (NJ)
Donald Lumsden	Co-Director, Office of Assessment, Kean College (NJ)
Linda Rudolph	Associate Vice President, Austin Peay State University (TN)
Marcia Mentkowski	Director, Office of Assessment, Alverno College (WI)
Glen Rogers	Assistant Director, Office of Assessment, Alverno College (WI)
Candy Young	Professor of Political Science, Northeast Missouri State University
Dary Erwin	Director, Office of Assessment, James Madison University (VA)
Trudy Banta	Director, Assessment Resource Center, UTK
Gary Pike	Associate Director, Assessment Resource Center, UTK
Stephen Dunbar	Professor, Lindquist Center, University of Iowa

APPENDIX F

Table of Contents of Assessment Resource Center Publications and List of Research Reports

ARTICLES ON ASSESSMENT AT UTK

Profiles of the UTK Assessment Program

Center for Assessment Research and Development
University of Tennessee, Knoxville
Knoxville, TN 37996-4350

TABLE OF CONTENTS

Page

Banta, T. W. (1986). Comprehensive program evaluation at the University of Tennessee, Knoxville: A response to changes in state funding policy. In T. W. Banta (Ed.), Performance funding in higher education: A critical analysis of Tennessee's experience. (pp. 37-50). Boulder, CO: National Center for Higher Education Management Systems. 1

Fisher, H. S. (1986). Planning and resource allocation at the University of Tennessee, Knoxville: Institutional responses to changes in state funding policies. In T. W. Banta (Ed.), Performance funding in higher education: A critical analysis of Tennessee's experience. (pp. 27-36). Boulder, CO: National Center for Higher Education Management Systems. 15

Banta, T. W. (1985). Use of outcomes information at the University of Tennessee, Knoxville. In P. T. Ewell (Ed.), Assessing educational outcomes (New directions for Institutional research, no. 47, pp. 19-32). San Francisco: Jossey-Bass. 25

Banta, T. W., & Moffett, M. S. (1987). Performance funding in Tennessee: Stimulus for program improvement. In D. F. Halpern (Ed.), Student outcomes assessment: What institutions stand to gain (New directions for higher education, no. 59, pp. 35-43). San Francisco: Jossey-Bass. 39

Banta, T. W., & Fisher, H. S. (1986). Putting a premium on results. Educational Record, 67, 54-59. 48

Banta, T. W., & Fisher, H. S. (1986, March 4). Measuring how much students have learned entails much more than simply testing then. Chronicle of Higher Education, 41-42. 53

ASSESSMENT BIBLIOGRAPHY

Selected Articles on Assessment and Educational Measurement

Center for Assessment Research and Development
University of Tennessee, Knoxville
Knoxville, TN 37996-4350

ASSESSMENT BIBLIOGRAPHY

Assessment Theory and Practice

Adelman, C. P. (Ed.). (1985). *Assessment in American higher education: Issues and Contexts*. Washington, D. C.: United States Government Printing Office.

Adelman, C. P. (Ed.). (in press). *Performance and judgment: Essays on principles and practice in higher education assessment*. Washington, D.C.: United States Government Printing Office.

Alexander, J. M., & Stark, J. S. (1987). *Focusing on student academic outcomes: A working paper*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Alverno College Faculty. (1979). *Assessment at Alverno College*. Milwaukee, WI: Alverno Productions.

American Association of State Colleges and Schools. (1986). *Proceedings from the conference on legislative action and assessment: Reason and reality*. Arlington, VA: Author.

Andrews, F. (1988, June). *The Tennessee model: Measuring educational outcomes in the College of Human Ecology*. Paper presented at the annual meeting of the American Home Economics Association, Baltimore, MD.

Association of American Colleges. (1985). *Integrity in the college curriculum: A report to the academic community*. Washington, D.C.: Author.

Astin, A. W. (1970a). The methodology of research on college impact, part one. *Sociology of Education*, *22*, 223-254.

Astin, A. W. (1970b). The methodology of research on college impact, part two. *Sociology of Education*, *22*, 437-450.

Astin, A. W. (1977). *Four critical years: Effects of college on beliefs, attitudes, and knowledge*. San Francisco: Jossey-Bass.

BIBLIOGRAPHY OF ASSESSMENT INSTRUMENTS

Information on Selected Educational Outcomes Measures

Center for Assessment Research and Development
University of Tennessee, Knoxville
Knoxville, TN 37996-4350

Table of Contents

Page

General Overview of Assessment Instruments 1

Assessment of General Education 2

Academic Profile 3

ACT Assessment Program 3

College BASE 4

Collegiate Assessment of Academic Proficiency 5

CLEP Education Assessment Series 5

CLEP General Education Examinations 6

College Outcome Measures Project 6

GRE: General Examinations 7

Assessment of Basic Skills 8

Descriptive Tests of Language Skills 8

Descriptive Tests of Mathematics Skills 9

New Jersey College Basic Skills Placement Tests 10

Test of Standard Written English 10

Assessment of Cognitive Development 11

Analysis of Argument 11

Erwin Scale of Intellectual Development 12

Measure of Epistemological Reflection 12

Reflective Judgment Interview 13

Test of Thematic Analysis 13

Watson-Glaser Critical Thinking Appraisal 14

Page

Assessment in the Major 15

American Assembly of Collegiate Schools of Business: Outcome Measurement Project 15

American Institute of Certified Public Accountants: Level II Achievement Test 16

GRE: Subject Tests 16

Major Field Achievement Tests 17

NTE Core Battery Tests 17

Assessment of Values 18

Defining Issues Test 18

Humanitarian/Civic Involvement Values 19

Kohlberg's Measure of Moral Development 19

Rokeach Value Survey 20

Assessing Involvement, Persistence, and Satisfaction 20

College Student Experiences Questionnaire 20

Program Self-Assessment Service 21

Student Opinion Survey 21

Student Reactions to College 22

Alverno College Assessment Instruments 22

Faculty-Designed Instruments: General Education 23

Faculty-Designed Instruments: Major Fields 24

Alumnae Questionnaire 25

Student as Learner Inventory 26

MATERIAL ON
LOCALLY-DEVELOPED TESTS

Articles and Working Papers on Test Development in the Major at UTK

Center for Assessment Research and Development
University of Tennessee, Knoxville
Knoxville, TN 37996-4350

TABLE OF CONTENTS

Page

Banta, T. W., & Schneider, J. A. (1988). Using faculty-developed exit exams to evaluate academic programs. Journal of Higher Education, 59, 69-83. 1

Learning Research Center. (1988). Comprehensive exams in major fields. Unpublished report, University of Tennessee, Knoxville. 9

Learning Research Center. (1987). Plan for development of comprehensive assessment procedures for seniors. Unpublished report, University of Tennessee, Knoxville. 12

Learning Research Center. (1987). Schedule for development of comprehensive assessment procedures for seniors. Unpublished report, University of Tennessee, Knoxville. 13

Sneed, J., & Carruth, B. R. (1988). Accountability in education: Assessing program outcomes. Hospitality Education and Research Journal, 15, 33-41. 14

SATISFACTION SURVEYS USED AT UTK

Copies of Locally-Developed Surveys Used at UTK

Center for Assessment Research and Development
University of Tennessee, Knoxville
Knoxville, TN 37996-4350

TABLE OF CONTENTS

Page

Van Liere, K. D., & Lyons, W. (1986). Measuring perceived program quality. In T. W. Banta (Ed.), Performance funding in higher education: A critical analysis of Tennessee's experience. (pp. 85-94). Boulder, CO: National Center for Higher Education Management Systems. 1

University of Tennessee - Knoxville. (1988). Student satisfaction with university programs and services. [Undergraduates]. Knoxville, TN: Author. 11

University of Tennessee - Knoxville. (1988). Survey of graduates. [Alumni]. Knoxville, TN: Author. 19

University of Tennessee - Knoxville. (1987). The student experience at UTK - a look back. [Dropouts]. Knoxville, TN: Author. 27

University of Tennessee - Knoxville. (1987). The graduate experience at UTK. [Graduate Students]. Knoxville, TN: Author. 35

RESEARCH PAPERS ON ASSESSMENT

Four Research Reports Prepared by the Center for Assessment Research and Development

Center for Assessment Research and Development
University of Tennessee, Knoxville
Knoxville, TN 37996-4350

TABLE OF CONTENTS

Page

Pike, G. R. (1988). Appendix A: Procedures for calculating test score improvement using nonequivalent groups. In T. W. Banta, 1987-88 performance funding report for the University of Tennessee, Knoxville. Learning Research Center, University of Tennessee, Knoxville. 1

Pike, G. R. (1988). Appendix B: A comparison of the College Outcome Measures Program (COMP) exam and the ETS Academic Profile. In T. W. Banta, 1987-88 performance funding report for the University of Tennessee, Knoxville. Learning Research Center, University of Tennessee, Knoxville. 10

Pike, G. R. (in review). Background characteristics, college experiences, and the ACT-COMP exam: Using construct validity to evaluate assessment instruments. Review of Higher Education. 28

Pike, G. R., & Phillippi, R. H. (1988, November). Relationships between self-reported coursework and performance on the ACT-COMP exam: An analysis of the generalizability of the differential coursework methodology. Paper presented at the annual meeting of the Association for the Study of Higher Education, St. Louis, Missouri. 72

STRATEGIES FOR ASSESSING OUTCOMES

November 6-7, 1989
Knoxville, Tennessee

Center for Assessment Research & Development
University of Tennessee, Knoxville
Knoxville, Tennessee 37996-4350

TABLE OF CONTENTS

Page

Alverno College

Assessment Program at Alverno College 1

Jacqueline Paskow (Ed.). Alverno College. Assessment Programs and Projects: A Directory. Washington, D.C.: AAHE Assessment Forum, 1988 2

Marcia Mentkowski and Georgine Loacker. Assessing and Validating the Outcomes of College. In Peter T. Ewell (Ed.), Assessing Educational Outcomes. New Directions for Institutional Research, no. 47. San Francisco: Jossey-Bass, 1985 5

Alverno Student Catalog. Introduction to Alverno Assessment 14

Georgine Loacker et al. Communication Assessment for Entering Students. Analysis and Communication at Alverno: An Approach to Critical Thinking. Milwaukee: Alverno Productions, 1984 15

Alverno Faculty. Alverno College: Example of a Comprehensive Performance Assessment for General Education. Assessment at Alverno College. Milwaukee: Alverno Productions, 1985 17

Marcia Mentkowski and Austin Doherty. Abilities that Last a Lifetime. AAHE Bulletin 34 (February 1984): 5-14 22

Alverno Faculty. Alverno College: Criteria for Faculty-Designed Instruments. Assessment at Alverno College. Milwaukee: Alverno Productions, 1985 28

Georgine Loacker and Patricia Jensen. The Power of Performance in Developing Problem Solving and Self Assessment Abilities. Assessment and Evaluation in Higher Education 13 (Summer 1988): 128-150 30

Georgine Loacker. Faculty as a Force to Improve Instruction Through Assessment. In J. H. McMillan (Ed.), Assessing Student Learning. New Directions for Teaching and Learning, no. 34. San Francisco: Jossey-Bass, 1988 53

Bibliography of Materials on Assessment at Alverno College 61

Page

Miami-Dade Community College

Jacqueline Paskow (Ed.). Miami-Dade Community College. *Assessment Programs and Projects: A Directory*. Washington, D.C.: AAHE Assessment Forum, 1988 66

Robert H. McCabe. Forward. *Computerized Adaptive Testing: The State of the Art at Three Community Colleges*. Laguna Hills, CA: League for Innovation in the Community College, 1988 68

Richard Schinoff and Larry Steed. The Computerized Adaptive Testing Program at Miami-Dade Community College, South Campus. *Computerized Adaptive Testing: The State of the Art at Three Community Colleges*. Laguna Hills, CA: League for Innovation in the Community College, 1988 69

Richard B. Schinoff. The Florida Plan: Mandated Entry and Exit Testing. Unpublished manuscript, Miami-Dade Community College 81

John Losak and Cathy Morris. Comparing Treatment Effects for Students who Successfully Complete College Preparatory Work. Research Report, Miami-Dade Community College, 1985. (ERIC Document Reproduction Service No. ED 264 921) 94

Ron Vorp. Completion of College Preparatory Requirements at Miami-Dade Community College by First-Time-In-College Students Entering Fall 1985. Research Report, Miami-Dade Community College, 1987 107

Northeast Missouri State University

Jacqueline Paskow (Ed.). Northeast Missouri State University. *Assessment Programs and Projects: A Directory*. Washington, D.C.: AAHE Assessment Forum, 1988 125

Abbie Ziffren. Northeast Missouri State University. *Liberal Education*. Washington, D.C.: Association of American College, 1988 129

Darrell Krueger and Margarita Heisserer. Assessment and Involvement: Investments to Enhance Learning. In Diane F. Halpern (Ed.), *Student Outcomes Assessment: What Institutions Stand to Gain*. New Directions for Higher Education, no. 59. San Francisco: Jossey-Bass, 1987 131

Page

University of Tennessee, Knoxville

Jacqueline Paskow. (Ed.). University of Tennessee, Knoxville. Assessment Programs and Projects: A Directory. Washington, D.C.: AAHE Assessment Forum, 1988 138

Trudy Banta and Homer Fisher. Putting a Premium on Results. Educational Record (Spring-Summer 1986): 54-58 140

Trudy Banta and Homer Fisher. Measuring How Much Students Have Learned Entails Much More than Simply Testing Them. Chronicle of Higher Education, march 4, 1987 144

Trudy Banta and Janet Schneider. Using Faculty-Developed Exit Examinations to Evaluate Academic Programs. Journal of Higher Education 59 (January-February 1988): 69-83 145

UTK Testing Statement 153

Tests in the Major Fields at UTK 154

Plan for Comprehensive Test Development 157

Schedule for Comprehensive Test Development 158

Kent Van Liere and William Lyons. Measuring Perceived Program Quality. In Trudy Banta (Ed.), Performance Funding in Higher Education. Boulder, CO: NCHEMS, 1985 159

UTK Student Satisfaction Survey 168

UTK Instructional Evaluation Report (Excerpt) 176

UTK Planning Goals 183

ASSESSMENT OF STUDENT OUTCOMES IN HIGHER EDUCATION

- Research Reports -
by
Faculty
at the
University of Tennessee, Knoxville

1988-89

Projects Sponsored by
The Office of the Chancellor, Academic Programs
and
The Center for Assessment Research and Development
1819 Andy Holt Avenue
Knoxville, Tennessee 37996-4350
(615) 974-2350

ASSESSMENT OF STUDENT OUTCOMES IN HIGHER EDUCATION

- Research Reports -
by
Faculty
at the
University of Tennessee, Knoxville

1988-89

Projects Sponsored by
The Office of the Chancellor, Academic Programs
and
The Center for Assessment Research and Development
1819 Andy Holt Avenue
Knoxville, Tennessee 37996-4350
(615) 974-2350

CONTENTS

Foreword i

Project Abstracts ii

I. Relationships Among COMP Examination Scores and Academic Achievement Variables for Nursing Students
Sandra P. Thomas and John Jozwiak
College of Nursing

II. Evaluation of Engineering Fundamentals Examination Results
Richard M. Bennett
Department of Civil Engineering
College of Engineering

III. Assessment Tests in Journalism and Public Relations
Edward Caudill
Susan Caudill
Paul Ashdown
School of Journalism
College of Communications

IV. The Relationship of College Outcomes to Post-Graduation Success
Michael K. Smith
Julie Thomas
Department of Psychology
College of Liberal Arts

MATERIALS ON THE TOPIC OF ASSESSING STUDENT OUTCOMES IN HIGHER EDUCATION

Available from:

Center for Assessment Research and Development
1819 Andy Holt Avenue
Knoxville, Tennessee 37996-4350
(615) 974-0883

CENTER FOR ASSESSMENT RESEARCH AND DEVELOPMENT
-- RESEARCH REPORTS --

No.	Title
88-01	Evidence from the Instructional Evaluation Program of UTK Effectiveness in Addressing Seven Principles for Good Practice in Undergraduate Education
88-02	Tennessee's Performance Funding Policy at Age 8
88-03	Students' Background Characteristics, Educational Experiences and Educational Outcomes: A Model for Evaluating Assessment Instruments
88-04	Selected Comparisons of the ACT COMP Exam and the ETS Academic Profile at the University of Tennessee, Knoxville
88-05	Procedures for Calculating Test Score Improvement Using Nonequivalent Groups
88-06	A Comparison of the College Outcome Measures Program (COMP) Exam and the ETS Academic Profile
88-07	Methods for Comparing Outcomes Assessment Instruments
88-08	Generalizability of the Differential Coursework Methodology: Relationships Between Self-Reported Coursework and Performance on the ACT-COMP Exam
89-01	Analysis of the Impact of Social Sciences Coursework on ACT-COMP Scores
89-02	A Comparison of Reliability and Difficulty Levels for Three Forms of the COMP Exam
89-03	A Collection of Assessment-Related Activities of the Faculty and Administration of the University of Tennessee, Knoxville
89-04	Campus-Based Assessment Research: Examples of the Genre
89-05	Utilizing Research on Postsecondary Student Outcomes: Evidence of Effectiveness
89-06	Using Construct Validity to Evaluate Assessment Instruments: A Comparison of the ACT-COMP Exam and the ETS Academic Profile

No.	Title
89-07	The Performance-Satisfaction Relationship Revisited: Specification and Testing of a Theoretical Model
89-08	Questions Faculty Ask About Assessment
89-09	Using Generalizability Theory in Institutional Research
89-10	A Comparison of the College Outcome Measures Program (COMP) and the Collegiate Assessment of Academic Proficiency (CAAP) Exams
89-11	The Performance of Black and White Students on the ACT-COMP Exam: An Analysis of Differential Item Functioning Using Samejima's Graded Model

APPENDIX G

Resource Appendices
from Performance and Judgment

PERFORMANCE AND JUDGMENT:

Essays on Principles and Practice in the Assessment of College Student Learning

Clifford Adelman
Mark Appelbaum
Leonard Baird
William Byham
John Centra
Stephen Dunbar
Sandra Graham
Jerilee Grandy
Jason Millman

Resource Appendices by Gary Pike

Edited by Clifford Adelman

Office of Research

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Appendix A:

An Annotated Bibliography on the Assessment of Student Educational Outcomes

by Gary Pike

The 1970s and 1980s have witnessed a dramatic growth in the literature on the assessment of student educational outcomes. Because of the variety of information available, this bibliography is not intended to be exhaustive. Instead, this review is provided as a starting point for the study of assessment. For convenience, the literature on assessment is organized around four themes: the basic principles underlying assessment programs, the identification of educational outcomes, the measurement of these outcomes and the analysis of outcomes data. Where an ED number is indicated (in brackets), the document cited is available through the ERIC Document Reproduction Service, 3900 Willer Ave., Alexandria, Va. 22304.

Principles Underlying Assessment Programs

Adelman, C., "To Imagine an Adverb: Concluding Notes to Adversaries and Enthusiasts," in Adelman, C. (ed.), Assessment in American Higher Education: Issues and Contexts. Washington, D.C.: U.S. Government Printing Office, 1986, pp. 73-82. Assessment (evaluation) is rooted in the nature of language, and because evaluation inheres in language, assessment shapes and is shaped by social and economic institutions. Based on this perspective, the author identifies several important measurement, organizational and policy concerns related to the growing interest in assessment by institutions of higher education.

Baker, E.L., "Critical Validity Issues in the Methodology of Higher Education Assessment," Assessing the Outcomes of Higher Education: Proceedings of the 1986 ETS Invitational Conference. Princeton: ETS, 1987, pp. 39-46. Baker examines the growing interest in assessment, arguing that assessment programs designed to measure effectiveness criteria established by State governments and regional accrediting associations often do not represent valid means of evaluating educational quality. The author contends that students' classroom experiences represent the best indicators of quality. As a result, the author recommends that assessment programs focus on outcomes directly related to classroom experiences.

Bergquist, W.H. and Armstrong, J.L. Planning Effectively for Educational Quality: An Outcomes-Based Approach for Colleges Committed to Excellence. San Francisco: Jossey-Bass, 1986. The authors provide a model for improving educational quality based on an examination of institutional mission. They advocate that institutions examine their missions, develop pilot programs designed to assist in accomplishing those missions, assess the effectiveness of the pilot programs, and then implement large-scale programs. Of relevance to those interested in assessment, the authors stress the importance of incorporating outcomes data in the planning process.

Chandler, J.W., "The Why, What, and Who of Assessment: The College Perspective," Assessing the Outcomes of Higher Education: Proceedings of the 1986 ETS Invitational Conference. Princeton: ETS, 1987, pp. 11-18. According to Chandler, assessment should focus on programs, not individuals. Assessment also should reflect the unique characteristics of an institution. Tailoring an assessment program to an institution encourages faculty ownership of the assessment program. In addition, the author explains why assessment should not be equated with testing.

Cross, K.P., "Using Assessment to Improve Instruction," Assessing the Outcomes of Higher Education: Proceedings of the 1986 ETS Invitational Conference. Princeton: ETS, 1987, pp. 63-70. The author argues that evaluations of classroom teaching should be an integral part of an assessment program. The author notes that one of the best ways to overcome faculty resistance to an assessment program is to provide the faculty with the tools to assess student learning and satisfaction.

Enthoven, A.C., "Measures of the Outputs of Education: Some Practical Suggestions for Their Development and Use," in Lawrence, B., Weathersby, G., and Patterson, V.W. (eds.), Outputs of Higher Education: Their Identification, Measurement, and Evaluation. Boulder, CO: Western Interstate Commission for Higher Education, 1970, pp. 51-60. [ED#043-296] The author makes three recommendations about the assessment of educational outcomes: first, assessment should be coupled with financial incentives; second, external evaluations (rather than course examinations) should be used in the assessment program; and third, assessment activities should be conducted by a central office of program analysis and review.

Ewell, P.T., The Self-Regarding Institution: Information for Excellence. Boulder, CO: National Center for Higher Education Management Systems, 1984. This volume focuses on the rationale underlying the assessment of student outcomes. The author begins by identifying four dimensions of student outcomes: "Knowledge Outcomes," "Skills Outcomes," "Attitude and Value Outcomes," and "Relationships with Society and with Particular Constituencies." The author also provides examples of how institutions have utilized outcomes data in their planning processes to improve education programs.

Heywood, J. *Assessment in Higher Education*. New York: John Wiley, 1977. The author provides a general overview of assessment, focusing on two critical aspects of educational improvement: the specification of objectives and the measurement of the extent to which these objectives are being met. The author concludes that educational improvement will occur only if assessment is made an integral part of a process of curriculum development and evaluation. The argument is based primarily on assessment practices in the United Kingdom, but the principles are universally applicable. A second edition is scheduled for publication in 1988.

Loacker, G., Cromwell, L., and O'Brien, K., "Assessment in Higher Education: To Serve the Learner," in Adelman, C. (ed.), *Assessment in American Higher Education: Issues and Contexts*. Washington, D.C.: U.S. Government Printing Office, 1986, pp. 47-63. Loacker *et al* assert that the ultimate goal of an assessment program should be to promote student learning and development. The authors thus view assessment as a multitrait and multimethod technique, and stress the need to develop evaluation activities outside the traditional student-faculty process.

Manning, T.E., "The Why, What, and Who of Assessment: The Accrediting Association Perspective," in *Assessing the Outcomes of Higher Education: Proceedings of the 1986 ETS Invitational Conference*. Princeton: ETS, 1987, pp. 31-38. Manning examines assessment from the perspective of the regional accrediting associations, noting that these associations have been advocating the use of assessment to improve institutional quality for several years.

Identification of Educational Outcomes

Before developing an assessment program, institutions must identify the outcomes to be assessed. In an effort to bring some coherence to this undertaking, several scholars have developed typologies of educational outcomes. While these typologies differ in many important respects, they all assume that student outcomes are multidimensional. The common outcomes described in these typologies can be grouped into four categories: cognitive outcomes (both knowledge and skills); affective outcomes (such as self-concept and moral development); attitudinal outcomes (including involvement and satisfaction); and outcomes expressed in terms of longer-term economic and social status (and, sometimes, participation in cultural, community and political life).

Alexander, J.M. and Stark, J.S., *Focusing on Student Academic Outcomes: A Working Paper*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning, 1987. These authors provide an overview of three typologies of student educational outcomes (Astin, Panos, and Creager). In addition, they provide brief descriptions of instruments designed to assess student outcomes in three areas: academic-cognitive, academic-motivational, and academic-behavioral.

Astin, A.W., "Measuring Student Outputs in Higher Education," in Lawrence, B., Weathersby, G., and Patterson, V.W. (eds.), Outputs of Higher Education: Their Identification, Measurement, and Evaluation. Boulder, CO: Western Interstate Commission for Higher Education, 1970, pp.75-84. [ED#043-296] Astin discusses the measurement and analysis of educational outputs from a modeling perspective, presenting a classic model of the educational process consisting of three components: student inputs, the college environment, and student outputs. In addition, the author stresses the importance of conducting multitrait/multimethod research over time.

Astin, A.W., "The Methodology of Research on College Impact, Part One." Sociology of Education, vol. 43 (1970), pp. 223-254. Arguing that research on student development should consist of multi-institutional longitudinal studies, Astin identifies several research designs and statistical procedures that are appropriate for assessing student educational outcomes. Astin also discusses technical issues related to detecting interaction effects and controlling for the effects of measurement error.

Astin, A.W., "Measurement and Determinants of the Outputs of Higher Education," in Solmon, L. and Taubman, P. (eds.), Does College Matter? Some Evidence of the Impacts of Higher Education. New York: Academic Press, 1973, pp. 107-127. In this article, Astin discusses the relationships between types of outcome, data, and time. The first two dimensions form a taxonomy consisting of cognitive-psychological, cognitive-behavioral, affective-psychological, and affective-behavioral outcomes.

Bloom, B.S. (ed.) Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain. New York: David McKay, 1956. This classic work details a hierarchy of educational objectives, ranging from lower-order outcomes, such as knowledge recall, to higher-order outcomes, such as synthesis and evaluation. Examples of measurement techniques for evaluating the attainment of each level in the hierarchy are also provided.

Brown, D.G., "A Scheme for Measuring the Output of Higher Education," in Lawrence, B., Weathersby, G., and Patterson, V.W. (eds.) Outputs of Higher Education: Their Identification, Measurement, and Evaluation. Boulder, CO: Western Interstate Commission for Higher Education, 1970, pp. 27-40. [ED#043-296] Brown examines the outputs of higher education from a measurement perspective, identifying five categories of educational outcomes and six characteristics of effective measurement. Based on his categories, several specific measures are identified. He also presents a simple model that can be used to assess educational outcomes.

College Outcomes Evaluation Program, New Jersey Department of Higher Education. "Final Report of the Student Learning Outcomes Subcommittee." Trenton: Author, 1987. In its report, the subcommittee examines the purpose of statewide assessment in New Jersey and identifies the types of student outcomes to be assessed. These outcomes include the general intellectual skills needed to analyze and utilize new information, the skills needed to understand and use different modes of inquiry, and the abilities necessary to appreciate various "continuities in the human experience."

Korn, H.A. *Psychological Models of the Impact of College on Students*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning, 1987. Korn describes five perspectives on the relationship between college experiences and student educational outcomes, and discusses the implications of recent advances in personality theory for the assessment of student outcomes. Korn also suggests several ways in which the models can be used to evaluate the impact of college on students.

Lenning, O.T. *Previous Attempts to Structure Educational Outcomes and Outcome-Related Concepts: A Compilation and Review of the Literature*. Boulder: National Center for Higher Education Management Systems, 1977. This report provides a taxonomy of educational outcomes based on two literature reviews. Impacts of higher education on individuals include intellectual development, emotional/cultural/social development, and physical development. In addition, the author includes potential impacts of higher education on society.

Pace, C.R., "Perspectives and Problems in Student Outcomes Research," in Ewell, P.T. (ed.), *Assessing Educational Outcomes*. New Directions for Institutional Research No. 47. San Francisco: Jossey-Bass, 1985, pp. 7-18. Pace presents a general overview of basic assessment techniques and instruments, identifying four categories of outcomes, as well as instruments designed to measure these outcomes. Given the variety of outcomes and instruments that may be used in an assessment program, the author stresses the importance of selecting outcomes consistent with the institution's mission and goals.

Pascarella, E.T., "College Environmental Influences on Learning and Cognitive Development: A Critical Review and Synthesis," in Smart, J.C. (ed.), *Higher Education: Handbook of Theory and Research*. New York Agathon Press, 1985, pp. 1-62. Pascarella presents a comprehensive synthesis of research on the factors influencing students' cognitive development during their college careers. He defines two categories of cognitive outcomes (knowledge and skills), discusses research relating to each, and identifies several instruments that have been used to measure various educational outcomes.

Measurement of Educational Outcomes

Development of Measures

Once relevant student outcomes have been identified, some method of measurement must be selected or developed for each outcome. Assessment programs generally have relied on two types of measures: surveys and tests. Several scholars have emphasized the importance of basing test and survey development on empirical research. For surveys, empirical research can be used to identify variables of interest and pilot tests can evaluate item quality. Scholars also have argued that test domains should be derived empirically and item analysis should be used to evaluate item quality.

Dillman, D.A. Mail and Telephone Surveys: The Total Design Method. New York: John Wiley, 1978. Dillman presents an overview of survey research methodology. Specific topics addressed include question writing and formatting, sampling, questionnaire administration, data analysis, and reporting of results. Of particular interest to assessment practitioners, is Dillman's approach to issues of development and administration from the perspective of maximizing response rates.

Dumont, R.G. and Troelstrup, R.L., "Exploring Relationships Between Objective and Subjective Measures of Instructional Outcomes." Research in Higher Education, vol. 12 (1980), pp. 37-51. This article reports research designed to identify the relationship between test scores and self-reports of learning. The authors found that the two indicators evidence moderate positive correlations, and conclude that self-reports are valid measures of learning.

Ebel, R.L., "Content Standard Test Scores," Educational and Psychological Measurement, vol. 22, (1962), pp. 15-25. This author recommends that test scores be interpreted as content standard scores, indicating a student's level of mastery of a given content area. Ebel argues that content scores should be used to supplement normative scores, and provides an extended example of the derivation of content standard scores using the Preliminary Scholastic Aptitude Test (PSAT).

Frederiksen, N. and Ward, W.C. Development of Measures for the Study of Creativity. GRE Research Report GREB 72-2P. Princeton:

Educational Testing Service, 1975. In the research described in this report, four tests of scientific creativity were developed: formulating hypotheses, evaluating proposals, solving methodological problems, and measuring constructs. Results from a universe of 4,000 students applying to graduate schools indicated that the measures evidence acceptable levels of reliability. In addition, scores on each of the four measures were found to be independent of scores on aptitude and achievement tests.

Gronlund, N.E. *Constructing Achievement Tests*. 3rd edition. Englewood Cliffs, NJ: Prentice-Hall, 1983. This short book provides a basic introduction to the construction of achievement tests. The author addresses all phases of test preparation and evaluation, and discusses issues related to the construction and scoring of both objective and essay tests.

Grosof, M.S. and Sardy, H. "Procedure: Measurement, Instrumentation, and Data Collection," in *A Research Primer for the Social and Behavioral Sciences*. Orlando, FL: Academic Press, 1985, pp. 133-168. These authors provide an overview of several measurement techniques, including surveys. They identify the various types of questions used in survey research and describe several approaches to scaling. They also provide several basic recommendations regarding question wording and discuss approaches to evaluating questionnaire reliability and validity.

Hambleton, R.K., "Determining Test Length," in Berk, R. A. (ed.), *A Guide to Criterion-Referenced Test Construction*. Baltimore: Johns Hopkins University Press, 1984, pp. 144-168. Hambleton notes that test length has important implications for the reliability and validity of criterion-referenced tests. Five different methods of determining test length are described, and factors influencing the selection of one of these methods are identified.

Marshall, J.C. and Hales, L.W. *Essentials of Testing*. Reading, MA: Addison-Wesley, 1972. Marshall and Hales provide a nontechnical discussion of a variety of approaches to test construction. In addition to identifying several principles of educational measurement, the authors detail the strengths and weaknesses of essay tests, completion tests, multiple-choice tests, and true-false tests.

Martuza, V.R. *Applying Norm-Referenced and Criterion-Referenced Measurement in Education*. Boston: Allyn and Bacon, 1977. Martuza describes the use of norm-referenced and criterion-referenced tests in educational research. Regarding norm-referenced tests, Martuza explains the importance of selecting appropriate norm groups, provides criteria for evaluating norms, and provides a step-by-step guide for test construction. Martuza also suggests several approaches to constructing criterion-referenced exams, including linguistic transformation, item-form/item-frame, amplified objectives, and facet design.

Mehrens, W.A. and Ebel, R.L. "Some Comments on Criterion-Referenced and Norm Referenced Achievement Tests." NCME Measurement in Education, vol. 10, (1979), pp. 1-8. [ED#182-324] The authors discuss two approaches to achievement testing: norm-referenced and criterion-referenced tests. In addition to defining these two types of tests, the authors conclude that norm-referenced tests are most appropriate for evaluating curriculum, while criterion-referenced exams are most appropriate for evaluating students' mastery levels.

Milton, O. and Eison, J.A. Textbook Tests: Guidelines for Item Writing. New York: Harper and Row, 1983. This is a basic introduction to writing test items. The authors underscore the importance of well-designed tests and offer several practical suggestions concerning item writing. They also include a series of exercises that allow the reader to identify the weaknesses of test questions.

Popham, W.J. "Specifying the Domain of Content or Behaviors," in Berk, R.A. (ed.), A Guide to Criterion-Referenced Test Construction. Baltimore: Johns Hopkins University Press, 1984, pp. 49-77. Popham addresses the issue of how to specify the areas of content and/or behavior to be covered in a test, stressing the importance of explicit test specification and congruent test item development. The author also makes several practical suggestions regarding the specification process that have implications for subsequent steps in the test development process.

Roid, G.H. "Generating the Test Items," in Berk, R.A. (ed.), A Guide to Criterion-Referenced Test Construction. Baltimore: Johns Hopkins University Press, 1984, pp. 49-77. Roid reviews several item-writing techniques and argues that the quality of the items generated in the test construction process can be enhanced if the items are based on empirical research. Four steps in the empirically derived item-writing process are identified.

Macro-Evaluation of Measures

Macro-evaluation of student outcomes measures is concerned with the reliability and validity of these measures. There are many approaches to evaluating the reliability of outcomes, ranging from classical correlational techniques to techniques that assess the internal consistency of measures based on generalizability theory. Because assessment efforts frequently have multiple purposes, multiple approaches to evaluating instrument reliability frequently are necessary.

The second major criterion for evaluating assessment instruments is validity. Instruments can be evaluated in terms of their content validity, criterion-related validity, and their construct validity. As with reliability, the type of validity evaluated may change depending on the purpose of the assessment program.

Anastasi, A. Psychological Testing. 4th edition. New York: Macmillan, 1976. This book is a basic reference work on the development, use, and evaluation of psychological tests. Topics addressed include ethical issues in the use of psychological tests, evaluation of instrument reliability and validity, and item analysis. In addition, the author identifies and analyzes several different types of tests, ranging from educational (achievement) tests to personality measures.

Berk, R.A. (ed.) A Guide to Criterion-Referenced Test Construction. Baltimore: Johns Hopkins University Press, 1984. This book contains essays that provide a technical discussion of the construction and evaluation of criterion-referenced tests. Essays on the evaluation of tests address issues of reliability and validity, noting that the decision to utilize a specific approach must be guided by the intended uses of the test data. In addition, essays on evaluating the reliability of cut-off scores and categorizations based on cut-off scores are included.

Cronbach, L.J. "Test Validation," in Thorndike, R.L. (ed.), Educational Measurement. 2nd edition. Washington, D.C.: American Council on Education, 1971, pp. 443-507. Cronbach's essay is a touchstone for understanding test validity. The author explains the goals of validation procedures and examines several types of validity: content validity, educational importance, construct validity, validity for selection, and validity for placement.

Cronbach, L.J. and Meehl, P.E. "Construct Validity in Psychological Tests." Psychological Bulletin, vol. 52, (1955), pp. 281-302. These authors examine procedures for validating psychological tests, focusing on construct validity. They indicate when construct validation of tests is appropriate and examine the assumptions underlying construct validity.

Gardner, E. "Some Aspects of the Use and Misuse of Standardized Aptitude and Achievement Tests," in Widgor, A.K. and Garner, W.R. (eds.), Ability Testing: Uses, Consequences, and Controversies: Part II. Washington, D.C.: National Academy Press, 1982, pp. 315-332. Gardner identifies six categories of misuse associated with an unquestioning reliance on standardized tests: acceptance of the test title for what the test measures; ignoring the error of measurement in test scores; use of a single test score for decision making; lack of understanding of test score reporting; attributing cause of behavior measured to the test; and test bias.

Linn, R. "Ability Testing: Individual Differences, Prediction, and Differential Prediction," in Widgor, A.K. and Garner, W.R. (eds.), Ability Testing: Uses, Consequences, and Controversies: Part II. Washington, D.C.: National Academy Press, 1982, pp. 335-388. This essay examines the use of standardized tests to assess individual differences. The author addresses issues related to criterion and predictive validity for educational and occupational performance, and the effects of socioeconomic and racial/ethnic differences.

Mehrens, W.A. and Lehmann, I.J. Using Standardized Tests in Education. 4th edition. New York: Longmann, 1987. Mehrens and Lehmann provide a general overview of measurement and evaluation in education. The chapter on reliability discusses approaches to estimating reliability based on correlational and generalizability theories. The chapter on validity identifies several different types of validity and presents methods for their estimation.

Stanley, J.C. "Reliability," in Thorndike, R.L. (ed.), Educational Measurement. 2nd edition. Washington, D.C.: American Council on Education, 1971, pp. 356-442. This basic reference on estimating reliability in educational measurement examines its topic in light of research on individual variation, and identifies sources of variation in test scores. The author also presents procedures for estimating reliability using classical correlational techniques and generalizability theory and discusses methods of estimating the reliability of change scores.

Wigdor, A.K. and Garner, W.R. (eds.) Ability Testing: Uses, Consequences, and Controversies: Part I. Washington, D.C.: National Academy Press, 1982. Part I of this work is the report of the Committee on Ability Testing of the Assembly of Behavioral and Social Sciences, National Research Council. The report provides an overview of ability testing (including the controversies associated with ability testing), identifies the uses of ability tests, and recommends a series of actions for the evaluation and improvement of ability tests.

Micro Evaluation of Measures

Micro-evaluation of outcomes measures is concerned with the analysis of individual questions (items). Several procedures are available to analyze test items, ranging from relatively simple item analysis procedures to mathematically sophisticated procedures based on Item Response Theory (IRT). Approaches based on IRT offer significant advantages (e.g., item difficulty estimates that vary according to the ability level of the student). IRT approaches also have important applications in detecting test item bias, equating test scores, and in developing tailored and computer-adaptive tests.

Berk, R.A. "Conducting the Item Analysis," in Berk, R.A. (ed.), A Guide to Criterion-Referenced Test Construction. Baltimore: Johns Hopkins University Press, 1984, pp. 97-143. Berk presents a technical discussion of the procedures that should be used to determine if individual test items function as they were intended. He emphasizes that both expert judgment and statistical techniques should be used to evaluate test items. In addition to providing a discussion of specific judgmental and statistical tests, he identifies step-by-step procedures for item analysis.

Diederick, P. Short-Cut Statistics for Teacher-Made Tests. Princeton: ETS, 1973. The author presents an introduction to the analysis of item quality for the less sophisticated mathematician. Topics addressed in the text include reliability, measurement error, and item analysis.

Hambleton, R.K. and Cook, L.L. "Latent Trait Models and Their Use in the Analysis of Educational Test Data." Journal of Educational Measurement, vol. 14, (1977), pp. 75-96. This article represents a general introduction to the use of latent trait (item response) models in education research. The authors begin by identifying the fundamental principles underlying latent trait theory, identify several common latent trait models, and suggest several applications for these models.

Hambleton, R.K. and Swaminathan, H. Item Response Theory: Principles and Applications. Boston: Kluwer-Nijhoff, 1985. The authors provide a basic reference work on item response theory. Topics addressed include ability scales, model fitting, and practical applications of item response theory.

Lord, F.M. Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980. In this technical discussion of item response theory. Lord identifies several applications of IRT, including tailored testing, ability testing, studies of item bias, and estimation of true-score distributions.

Office for Minority Education. An Approach for Identifying and Minimizing Bias in Standardized Tests: A Set of Guidelines. Princeton: ETS, 1980. This report explains the issues related to bias in testing, and presents a series of guidelines for eliminating item bias in test construction and evaluating existing tests to detect biased items.

Assessment of Writing/Using Essay Examinations

Breland, H.M., Camp, R., Jones, R.J., Morris, M.M., and Rock, D.A. Assessing Writing Skill. Research Monograph No. 11.

New York: College Entrance Examination Board, 1987. These authors describe a study designed to assess writing skill at six colleges and universities. Results indicated that the unreliability of essay scoring could be alleviated by relying on multiple essays or by combining objective and essay tests. The authors also demonstrate the use of a variety of data analysis techniques. Both essay and objective tests were found to be about equal in their predictive validity. The authors conclude that multi-method assessment techniques offer both theoretical and practical advantages over other approaches.

Coffman, W.E. "Essay Examinations," in Thorndike, R.L. (ed.), *Educational Measurement*. 2nd edition. Washington, D.C.: American Council on Education, 1971, pp. 271-302. In this chapter, the author examines the advantages and limitations of essay tests as assessment tools, with specific attention to issues related to the reliability and validity. In addition, the author offers several suggestions for improving the use of essay exams.

Coffman, W.E. "On the Validity of Essay Tests of Achievement." *Journal of Educational Measurement*, vol. 3, (1966), pp. 151-156. This author reports research concerning methods of validating essay and objective tests. Traditionally, essay and objective tests have been correlated in order to demonstrate the predictive validity of objective tests. The author examines the predictive power of a sample of essay questions independent of objective measures.

Cooper, P.L. *The Assessment of Writing Ability: A Review of Research*. GRE Research Report GREB 82-15R. Princeton: ETS, 1984. The psychometric and practical issues related to the assessment of writing are the focus of this review. The author notes that although essay tests are considered to be more valid than multiple-choice tests, variability in subjects' scores may be influenced by a wide range of irrelevant factors. The author contends that when procedures to correct for threats to reliability and validity are employed, essay tests correlate very highly with multiple-choice tests.

Crocker, L. "Assessment of Writing Skills Through Essay Tests," in Bray, D. and Belcher, M. J. (eds.), *Issues in Student Assessment*. New Directions for Community Colleges, No. 59. San Francisco: Jossey-Bass, 1987. In discussing the use of essay tests in assessing basic writing skills, the author provides a rationale for using essay exams to assess writing abilities and identifies the steps required to develop a writing assessment program. These steps include: developing prompts (topics), developing scoring procedures, training raters, field testing, and administering the instruments. The author also examines issues related to the reliability and validity of essay exams.

Keeley, S.M., Browne, N.M., and Kreutzer, J.S. "A Comparison of Freshmen and Seniors on General and Specific Essay Tests of Critical Thinking." *Research in Higher Education*, vol. 17, (1982), pp. 139-154. These authors report research utilizing essay tests to evaluate the critical-thinking skills of freshmen and seniors. Results indicate that educational experiences produce significant gains in critical-thinking skills. An important finding for assessment practitioners was that significant differences in students' writing samples are related to the type of instructions (general or specific) provided for the assessment.

Steele, J.M. "The Assessment of Writing Proficiency via Qualitative Ratings of Writing Samples." Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, 1979. [ED#175-944] Steele examines several strategies for improving the reliability of raters' evaluations of writing samples. Research has indicated that increasing the number of writing samples per student to three significantly increases interrater reliability. However, using more than two raters does not improve reliability significantly.

Steele, J.M. "Trends and Patterns in Writing Assessment." Paper presented at the Annual Conference on the Assessment of Writing." San Francisco, 1985. [ED#268-146] The author describes the writing assessment portion of the College Outcome Measures Project (COMP) Composite Examination. He notes that the COMP exam, unlike many writing assessment instruments, focuses on writing in problem solving and critical thinking situations. Instead of providing a single holistic rating, the COMP writing assessment provides scores in three areas of writing proficiency.

White, E.M. *Testing and Assessing Writing*. San Francisco: Jossey-Bass, 1985. This book offers an overview of issues related to the assessment of writing. Included are discussions of holistic scoring, the use of proficiency tests, selection and/or development of writing tests, and the evaluation/scoring of writing assignments.

Nontraditional Outcomes Measures

During the last decade, there has been a marked increase in the use of nontraditional approaches to assess student educational outcomes. As a general rule, these approaches have been intended as supplements to existing measurement techniques. Reliance on multiple assessment methods has been shown to improve the validity of evaluations.

Most of the nontraditional measurement approaches have focused on the assessment of student performance through such techniques as assessment centers, simulations, and external evaluators. Exceptions have included computer-adaptive testing methods and the use of unobtrusive (nonreactive) measures to gather assessment data.

Berk, R.A. (ed.) Performance Assessment: Methods and Applications. Baltimore: Johns Hopkins University Press, 1986. This basic technical reference work on performance assessment includes essays covering a variety of performance assessment methods ranging from behavior rating scales to assessment center techniques. The authors also identify applications of performance assessment in business, medicine and the law, teaching, and the evaluation of communication skills.

Fong, B. The External Examiner Approach to Assessment. Washington, D.C.: AAHE Assessment Forum, 1987. This monograph provides an overview of the use of external examiners as an assessment tool. While both British and American experience is considered, special attention is paid to how American institutions are using external examiners to evaluate student mastery of content in courses and disciplines. The author also discusses issues of reliability and validity as they relate to the use of external examiners.

Hsu, T. and Sadock, S.F. Computer-Assisted Test Construction: The State of the Art. Princeton: ETS, 1985. [ED#272-515] These authors discuss both theory and applications of computers in developing and administering tests. They contend that adaptive testing is the one example of the successful use of computers to improve the quality of the assessment process.

Millman, J. "Individualizing Test Construction and Administration by Computer," in Berk, R.A. (ed.), A Guide to Criterion-Referenced Test Construction. Baltimore: Johns Hopkins University Press, 1984, pp. 78-96. Millman presents a technical review of the application of computers in test construction and administration. Specific topics include traditional attempts to individualize testing (equivalent forms of a test), item banking, and computer-adaptive testing. Millman notes that the proliferation of computer-adaptive tests has created a need for further research on the cost effectiveness of this approach. Millman concludes that assessment practitioners should be very cautious in utilizing computer-adaptive tests developed outside their own institutions.

Stillman, P.L. and Swanson, D.B. "Ensuring the Clinical Competence of Medical School Graduates Through Standardized Patients." Archives of Internal Medicine, vol. 147, (1987), pp. 1049-1052. These authors discuss the use of "Standardized Patients" to assess medical students' interviewing and physical examination skills. "Standardized Patients" are trained to function in multiple roles and to simulate a physician-patient encounter. Preliminary research suggests that this approach offers a realistic means of standardizing performance assessment for medical school graduates.

Terenzini, P.T. "The Case for Unobtrusive Measures," *Assessing the Outcomes of Higher Education: Proceedings of the 1986 ETS Invitational Conference*. Princeton: ETS, 1987, pp. 47-61. Terenzini argues that traditional data collection methods (tests, surveys, and interviews) should be supplemented by unobtrusive measurement techniques that can overcome the sources of measurement error present in other approaches, and are relatively inexpensive to administer. The author also presents a typology of unobtrusive measurement techniques that can be used to guide the selection of particular measures.

Urry, V.W. "Tailored Testing: A Successful Application of Latent Trait Theory." *Journal of Educational Measurement*, vol. 14, (1977), pp. 181-196. Urry describes the role of Item Response Theory in the development and administration of tailored (computer-adaptive) tests. In addition, he analyzes the computer-adaptive test used by the U.S. Civil Service Commission and identifies future uses for computer-adaptive ability tests.

Webb, E.J., Campbell, D.T., Schwartz, R.D., and Sechrest, L. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally, 1966. In this classic short work, the authors present several reasons for supplementing traditional measurement techniques with unobtrusive measures, and identify several approaches to unobtrusive measurement. These measurement techniques include physical traces, archival data, simple observations, and contrived observation.

Webb, E. and Weick, K.E. "Unobtrusive Measures in Organization Theory: A Reminder," in Maanen, J.V. (ed.), *Qualitative Methodology*. Beverly Hills, CA: Sage Publications, 1983, pp. 209-224. In this chapter, the authors examine the use of unobtrusive measures in organizational research. Of particular interest to assessment practitioners, the authors identify six ways in which unobtrusive measurement can modify traditional data collection methods.

Value-Added Analysis of Outcomes Data

While the concept of the value added by a college education is compelling, several scholars have criticized the concept when value added is defined as a gain or difference score. Writers have suggested several alternatives to simple gain, including residual and base-free measures of gain. Still other scholars have suggested that repeated measures designs be used in value-added analyses.

Bereiter, C. "Some Persisting Dilemmas in the Measurement of Change," in Harris, C.W. (ed.), *Problems in Measuring Change*. Madison: University of Wisconsin Press, 1963, pp. 3-20. This introduction to the problems inherent in the use of change scores identifies and analyzes three dilemmas associated with their use: over-correction/under-correction; unreliability/ invalidity; and physicalism/subjectivism.

Cronbach, L.J. and Furby, L. "How We Should Measure 'Change'--Or Should We?" *Psychological Bulletin*, vol. 74, (1974), pp. 68-80. See also "Errata." *Ibid*., p. 218. This technical discussion of methods for calculating change (difference) scores offers formulas for calculating "true" scores, "residual" scores, and "base-free" measures. Given the purposes for which change scores are used, the authors recommend a multivariate approach for evaluating change. Researchers interested in using the formulas presented in this article should carefully read the "Errata."

DuBois, P.H. "Correlational Analysis in Training Research," in DuBois, P.H. and Mayo, G.D. (eds.), *Research Strategies for Evaluating Training*. Chicago: Rand McNally, 1970. DuBois discusses the use of change scores in correlational research, explains the rationale underlying residual change scores, and presents a formula for calculating residual gain scores.

Gaito, J. and Wiley, D.E. "Univariate Analysis of Variance Procedures in the Measurement of Change," in Harris, C.W. (ed.), *Problems in Measuring Change*. Madison: University of Wisconsin Press, 1963, pp. 60-84. The authors present a basic description of the use of univariate analysis of variance to analyze repeated measures data. They begin by describing the assumptions underlying the univariate analysis of variance and then present a mathematical explanation of the univariate model. The authors also identify several procedures that may be used to minimize the effects of contaminating influences.

Horst, P. "Multivariate Models for Evaluating Change," in Harris, C.W. (ed.),*Problems in Measuring Change*. Madison: University of Wisconsin Press, 1963, pp. 104-121. Horst describes the theory underlying a general multivariate model for the evaluation of change. Horst examines the assumptions underlying a multivariate approach and provides a mathematical model for the multivariate analysis of change. This model relies on a multi-categorical matrix in which row vectors represent subjects and column vectors represent administrations of an instrument.

Lord, F.M. "Elementary Models for Measuring Change," in Harris, C.W. (ed.), *Problems in Measuring Change*. Madison: University of Wisconsin Press, 1963, pp. 21-38. Lord examines several problems inherent in the use of difference scores, including unreliability and regression effects. He notes that these problems can produce spurious relationships between gain scores and other variables;

and presents a method for calculating true gain scores.

McMillan, J.H., "Techniques for Evaluating Value-Added Data: Judging Validity, Improvement, and Causal Inferences." Paper presented at the annual meeting of the American Educational Research Association, Washington, 1987. The author identifies several limitations of value-added analyses and describes methods of overcoming these limitations. He suggests that researchers utilize appropriate research designs and statistical procedures when evaluating difference scores. One possible source of confirmatory data would be faculty judgments.

Pascarella, E.T., "Are Value-Added Analyses Valuable?" Assessing the Outcomes of Higher Education: Proceedings of the 1986 ETS Invitational Conference. Princeton: ETS, 1987, pp. 71-92. Pascarella presents a nontechnical discussion of the benefits and problems of relying on value-added data. He suggests several different methods of overcoming the problems associated with the use of difference scores and presents modeling techniques that can be used with value-added data.

Rogosa, D., Brandt, D., and Zimowski, M. "A Growth Curve Approach to the Measurement of Change." Psychological Bulletin, vol. 92, (1983), pp 726-748. These authors argue that the criticism of change scores as unreliable does not mean that they should be abandoned.

Tucker, L.R., Damarin, F., and Messick, S. "A Base-Free Measure of Change." Psychometrika, vol. 31, (1966), pp. 457-473. In this article, the authors identify and discuss problems with the calculation and use of simple gain scores. They recommend the use of a base-free measure of change, and provide formulas for calculating this measure.

Willett, J.B., "Questions and Answers in the Measurement of Change," in Rothkopf, E.R. (ed.), Review of Research in Education. volume 14. Washington, D.C.: American Educational Research Association, 1987. According to Willett, the measurement and analysis of growth (change) is central to evaluating educational effectiveness. Willett contends that the criticisms of growth measures that have been directed at two-wave (pre- and posttest) designs are overstated. Although Willett identifies instances in which simple difference scores can be reliable and valid, he recommends a multi-wave approach to measuring change.

Wolfle, L.M. "Applications of Causal Models in Higher Education," in Smart, J.C. (ed.), Higher Education: Handbook of Theory and Research. New York: Agathon Press, 1985. Wolfle examines the use of causal modeling as a research tool in the assessment of educational outcomes, explaining its assumptions and analyzing the concepts of causation and the decomposition of effects. The

author also discusses the specification of recursive and nonrecursive models, and the use of causal models with latent variables.

Appendix B:

Review of Assessment Instruments

by Gary Pike

This review is designed to provide brief descriptions of the technical characteristics of many of the instruments mentioned in this book. For convenience, the descriptions are organized around six types of outcomes: general education, basic skills, cognitive development, learning in the discipline, values, and motivation. Within each outcome area, tests are listed in alphabetical order.

Assessment of General Education

Academic Profile

Publisher: ETS College and University Programs, Educational Testing Service, Princeton, NJ 08541-0001; Scales: Total Score, Humanities, Social Science, Natural Sciences, Reading, Writing, Critical Thinking, and Mathematics; Length: 48-144 items; Time: 1-3 hours.

The Academic Profile has been developed by ETS and the College Board to assess the effectiveness of general education programs. The Academic Profile is available in two forms: a one-hour exam providing group feedback, and a three-hour exam providing individual feedback. A panel of experts in the content fields supervised test construction, assisting with questions of content validity. Because ETS is making the Academic Profile available for pilot testing during the 1987-1988 academic year, further information about the reliability and validity of this test is not available at this time.

ETS College and University Programs. The Academic Profile. Princeton: ETS, 1981.

ACT Assessment Program

Publisher: American College Testing Program, P.O. Box 168, Iowa City, IA 52240; Scales: Composite Score, English Usage, Mathematics Usage, Social Studies Reading, Natural Science Reading; Length: 40-75 items/test; Time: 30-50 minutes/test.

The ACT Assessment Program was developed as a series of college entrance and placement examinations for high school graduates. Depending on the coefficients used, reliability estimates have ranged from .73 to .91. Research has found that the ACT Assessment is capable of predicting subsequent performance in college, including cumulative grade point average and performance in specific classes. However, research at Tennessee Technological University could not demonstrate a relationship between gains on the ACT Assessment exam and students' experiences in college, raising questions about the validity of the ACT Assessment exam as a measure of educational effectiveness.

American College Testing Program. Assessing Students on the Way to College: Technical Report for the ACT Assessment Program. Iowa City, IA: ACT, 1973.

American College Testing Program. College Student Profiles: Norms for the ACT Assessment. Iowa City, IA: ACT, 1987.

Dumont, R.G. and Troelstrup, R.L. "Measures and Predictors of Educational Growth with Four Years of College." Research in Higher Education, vol. 14, (1981), pp. 31-47.

Munday, L.A. "Correlations Between ACT and Other Predictors of Academic Success in College." College and University, vol. 44, (1968), pp. 67-76.

Richards, J.M., Jr., Holland, J.L., and Lutz, S.W. "Prediction of Student Accomplishment in College." Journal of Educational Psychology, vol. 58, (1967), pp. 343-355.

College Basic Academic Subjects Examination

Publisher: Center for Educational Assessment, University of Missouri-Columbia, 403 South Sixth Street, Columbia, MO 65211; Scales: English, Mathematics (2), Science, Social Studies, Reading, Reasoning, and Writing (optional); Length: approximately 40-120 items; Time: 1-3 hours.

The College Basic Academic Subjects Examination (College BASE) is a criterion-referenced achievement test that can be used to evaluate individuals or programs. One-and three-hour forms of the exam are available. Content validity of the College BASE was achieved by using expert reviewers during the test construction process. Because the exam is being pilot tested during the 1987-88 academic year, additional information on reliability and validity has not been made available.

Center for Educational Assessment. College BASE. Columbia, MO: University of Missouri-Columbia, 1987.

Collegiate Assessment of Academic Proficiency

Publisher: American College Testing Program. 2201 N. Dodge St., P.O. Box 168, Iowa City, Iowa 52243; Scales: Reading, Mathematics, Writing, and Critical Thinking. Length: 175 items for all four modules in pilot administration plus 2 prompts for writing sample; Time: 40 minutes for each module and 40 minutes for the writing sample.

The Collegiate Assessment of Academic Proficiency (CAAP) is a new standardized test intended to assist institutions in evaluating their general education programs by assessing those academic skills typically developed during the first two years of college. The CAAP is available in modules, and institutions may add questions to the exam, thereby tailoring the exam to their curriculum. Because the exam is being pilot-tested beginning in 1988, information on reliability and validity is not available.

American College Testing Program. Collegiate Assessment of Academic Proficiency: Test Specifications and Sample Items. Iowa City, IA: ACT, 1988.

CLEP Education Assessment Series

Publisher: The College Board. 45 Columbus Ave. New York, NY 10023-6917. Scales: English Composition, Mathematics; Length: 40-45 questions per scale; Time: 45 minutes per module.

The Education Assessment Series (EAS) consists of two tests intended to provide comprehensive, nationally-normed data in a relatively short administration time and at low cost. Because multiple forms of the exams will be available, institutions may administer them twice and calculate the "value added" by general education. The tests are being piloted in 1988, hence information concerning reliability and validity is not yet available.

The College Board. CLEP Introduces the Education Assessment Series. New York: Author, 1988.

CLEP General Education Examinations

Publisher: College Entrance Examination Board, 45 Columbus Ave. New York, NY 10023-6917; Tests: English Composition, Humanities, Mathematics, Natural Science, and Social Science/History; Length: 55-150 items/test; Time: 90 minutes/test.

The College-Level Examination Program (CLEP) General Examinations cover five content areas and were designed to provide college credit for non-college learning. Reliabilities for the five tests range from .91 to .94. Using panels of experts in the content fields, the CLEP test development process has achieved satisfactory levels of content validity. While research has linked CLEP scores to performance in introductory college courses, no studies have been conducted on the validity of the CLEP exams as program evaluation instruments.

College Entrance Examination Board. Technical Manual Overview. Princeton: ETS, 1984.

College Entrance Examination Board. Outcomes Assessment in Higher Education. Princeton: ETS, 1986.

College Outcome Measures Project

Publisher: ACT, P.O. Box 168, Iowa City, IA 52243; Scales: Total Score, Functioning within Social Institutions, Using Science and Technology, Using the Arts, Communicating, Solving Problems, Clarifying Values, Writing (CE), Speaking (CE), Reasoning and Communicating (CE); Length: 60-99 items; Time: 2.5-4.5 hours.

The College Outcome Measures Project (COMP) examination was designed to measure the knowledge and skills necessary for effective functioning in adult society. This exam is available in two forms: the Objective Test (OT), consisting of 60 multiple-choice items; and the Composite Examination (CE), containing the same multiple-choice questions and speaking/writing exercises. Estimates of reliability for the COMP sub-scales were satisfactory (ranging from .63 to .81) although research on its validity as an assessment instrument has produced mixed results. Studies by ACT have shown that COMP scores are related to general education coursework and student involvement; however, other research by colleges themselves has failed to find a link between COMP scores (or gains on the COMP) and effective academic programs.

Banta, T.W., Lambert, E.W., Pike, G.R., Schmidhammer, J.L. and Schneider, J.A., "Estimated Student Score Gain on the ACT COMP Exam: Valid Tool for Institutional Assessment?" Paper presented at the annual meeting of the American Educational Research Association, Washington, 1987. [ED#281-892]

Forrest, A. Increasing Student Competence and Persistence: The Best Case for General Education. Iowa City, IA: ACT National Center for the Advancement of Educational Practices, 1982.

Forrest, A. and Steele, J.M. Defining and Measuring General Education Knowledge and Skills. Iowa City, IA: ACT, 1982.

Kitabchi, G. "Multivariate Analysis of Urban Community College Student Performance on the ACT College Outcomes Measures Program Test." Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1985. [ED#261-091]

Steele, J. M. "Assessing Speaking and Writing Proficiency via Samples of Behavior." Paper presented at the annual meeting of the Central States Speech Association, 1979. [ED#169-597]

Graduate Record Examinations Program: General Examinations

Publisher: Graduate Record Examinations Board, CN 6000, Princeton, NJ, 08541-6000; Scales: Verbal (antonyms, analogies, sentence completions, reading passages), Quantitative (quantitative comparisons, mathematics, data interpretation), Analytic (analytical reasoning, logical reasoning); Length: 50-76 items per sub-test; Time: 3 hours, 30 minutes.

The General Examinations of the GRE are nationally normed tests designed to assess learned abilities that are not related to any particular field of study, but that are related to the skills necessary for graduate study. Research on the GRE General Examinations has revealed high levels of reliability (.89 to .92) for the three tests. Reliability estimates for the nine item-types are somewhat lower (.60 to .90). Research has also found that test (and item-type) scores are related to undergraduate performance as well as to performance in graduate school.

Adelman, C. The Standardized Test Scores of College Graduates, 1964-1982. Washington, D.C.: U.S. Government Printing Office, 1984.

Conrad, L., Trismen, D., and Miller, R. Graduate Record Examinations Technical Manual. Princeton: ETS, 1977.

Fortna, R.O. Annotated Bibliography of the Graduate Record Examinations. Princeton: ERIC Clearinghouse on Tests, Measurement, and Evaluation, 1980.

Graduate Record Examinations Board. GRE Guide to the Use of the Graduate Record Examinations Program. Princeton: ETS, 1987.

Swinton, S.S. and Powers, D.E. A Study of the Effects of Special Preparation on GRE Analytical Scores and Item Types. GRE Research Report GREB 78-2R. Princeton: ETS, 1982.

Wilson, K.M. The Relationship of GRE General Test Item-Type Part Scores to Undergraduate Grades. GRE Research Report 81-22P. Princeton: ETS, 1985.

Assessment of Basic Skills

Descriptive Tests of Language Skills

Publisher: Descriptive Tests of Language Skills, Educational Testing Service, Mail Drop 22E, Princeton, NJ 08541; Scales: Reading Comprehension, Logical Relationships, Vocabulary, Usage, and Sentence Structure; Length: 30-50 items/test; Time: 15-30 minutes/test.

The Descriptive Tests of Language Skills (DTLS) consist of five tests designed for the placement of students in college English classes. These tests may be used separately or in combination. Because of their low difficulty levels, the DTLS are most appropriate for identifying students in need of remediation. Research by ETS indicates that all five tests evidence acceptable reliability (from .82 to .89); that the DTLS are correlated with writing ability and other measures of academic ability, such as ACT scores; and that performance on the DTLS predicts college grade point average. Studies have not examined the appropriateness of the DTLS as instruments for evaluating program quality.

College Entrance Examination Board. *Guide to the Use of the Descriptive Tests of Language Skills*. Princeton: ETS, 1985.

Snowman, J., Leitner, D.W., Snyder, V. and Lockhart, L., "A Comparison of the Predictive Validities of Selected Academic Tests of the American College Test (ACT) Assessment Program and the Descriptive Tests of Language Skills for College FreshmEn in a Basic Skills Program." *Educational and Psychological Measurement*, vol. 40, (1980), pp. 1159-1166.

Snyder, V. and Elmore, P.B., "The Predictive Validity of the Descriptive Tests of Language Skills for Developmental Students Over a Four-Year College Program." *Educational and Psychological Measurement*, vol. 43, (1983), pp. 1113-1122.

Descriptive Tests of Mathematics Skills

Publisher: Descriptive Tests of Mathematics Skills, Educational Testing Service, Mail Drop 22E, Princeton, NJ 08541; *Scales*: Arithmetic Skills, Elementary Algebra Skills, Intermediate Algebra Skills, and Functions and Graphs; *Length*: 30-35 items/test; *Time*: 30 minutes/test.

The four tests in the Descriptive Tests of Mathematics Skills (DMTS), used separately or in combination, are designed to assess mathematics skills for placement purposes. Because of their low item difficulty levels, these tests are not appropriate for differentiating among students with high levels of math skills. Research has indicated that the DTMS examinations evidence acceptable reliability (.84 to .91) and that the DTMS are related to measures of academic ability and performance in introductory math courses, particularly remedial courses.

Bridgeman, B., "Comparative Validity of the College Board Scholastic Aptitude Test--Mathematics and the Descriptive Tests of Mathematics Skills for Predicting Performance in College Mathematics Courses." *Educational and Psychological Measurement*, vol. 42, (1982), pp. 361-366.

College Entrance Examination Board. *Guide to the Use of the Descriptive tests of Mathematics Skills*. Princeton: ETS, 1985.

New Jersey College Basic Skills Placement Tests

Publisher: NJCBSPT, College Entrance Examination Board, Educational Testing Service, Princeton, NJ 08541; *Scales*: Writing, Reading Comprehension, Sentence Sense, Math Computation, Elementary Algebra, Composition (composite score), and Total English (composite score); *Length*: 168 items; *Time*: 3 hours.

The New Jersey College Basic Skills Placement Test (NJCBSPT) consists of five tests designed to meet the requirements of the assessment and evaluation program developed by the New Jersey Board of Higher Education. In addition to the five test scores, two composite scores can be derived from the language assessment parts of the test. Reliability estimates for the seven subscales range from .83 to .92. The content validity of the NJCBSPT was achieved by providing for constant review during test construction by a panel of experts from the New Jersey Basic Skills Council. Studies on the construct validity and predictive validity of the NJCBSPT are currently underway.

College Entrance Examination Board. *The New Jersey College Basic Skills Placement Test Program: Your Information Base for Outcomes Assessment*. Princeton: ETS, 1987.

Office of College Outcomes. *Appendices to the Report of the New Jersey Board of Higher Education from the Advisory Committee to the College Outcomes Evaluation Program*. Trenton, NJ: New Jersey Department of Higher Education, 1987.

Test of Standard Written English

Publisher: Test of Standard Written English, College Entrance Examination Board, Princeton, NJ 08541; *Scales*: Total Score; *Length*: 50 items; *Time*: 30 minutes.

The Test of Standard Written English (TSWE) is designed to measure a student's ability to use the language contained in most college textbooks. Research has found that the TSWE evidences acceptable reliability, and is predictive of performance in freshman English courses. The TSWE also has been found to be predictive of performance during the Junior year. Indeed, the TSWE has been found to be as good a predictor of performance as longer, more complex exams.

Bailey, R.L. "The Test of Standard Written English: Another Look." *Measurement and Evaluation in Guidance*, vol 10, (1977), pp. 70-74.

Michael, W.B. and Shaffer, P. "A Comparison of the Validity of the Test of Standard Written English (TSWE) and of the California State University and Colleges English Placement Test (CSUC-EPT) in the Prediction of Grades in a Basic English Composition Course and of Overall Freshman-Year Grade Point Average." Educational and Psychological Measurement, vol. 39, (1979), pp. 131-145.

Suddick, D.E. "A Re-examination of the Use of the Test of Standard Written English and Resulting Placement for Older Upper-Division and Master's Level Students." Educational and Psychological Measurement, vol. 42, (1982), pp. 367-369.

Suddick, D.E., "The Test of Standard Written English and Resulting Placement Patterns: A Follow-up of Performance of Older Upper-Division and Master Level Students." Educational and Psychological Measurement, vol. 41, (1981), pp 599-601.

Assessment of Cognitive Development

Analysis of Argument

Author: David G. Winter, Department of Psychology, Wesleyan University, Middletown, CT 06457; Scales: Total Score; Length: two exercises; Time: 10 minutes.

The Analysis of Argument is a production measure designed to assess clarity and flexibility of thinking skills. After reading a passage representing a particular position on a controversial issue, subjects are asked to write a response disagreeing with the original position. After 5 minutes, they are then instructed to write a short essay that agrees with the original position. The two essays are scored using a 10-category scheme. Because inter-rater agreement is a function of training, the authors do not provide estimates of reliability. The authors do report that studies have found that scores on the Analysis of Argument test are significantly related to other measures of cognitive development, as well as to previous educational experiences.

Stewart, A.J. and Winter, D.G. Analysis of Argument: An Empirically Derived Measure of Intellectual Flexibility. Boston: McBer and Company, 1977.

Erwin Scale of Intellectual Development

Author: T. Dary Erwin, Office of Student Assessment, James Madison University, Harrisonburg, VA 22801; Scales: Dualism, Relativism, Commitment, Empathy; Length: 86 items; Time: untimed.

The Erwin Scale of Intellectual Development (SID) was designed to measure intellectual development based on Perry's scheme, three of the four sub-scales (dualism, relativism and commitment) paralleling Perry's categories of intellectual development. Research on the SID has found that all four sub-scales evidence acceptable reliability (.70 to .81) and that the SID is significantly related to other measures of development, including measures of identity and involvement.

Erwin, T.D., "The Scale of Intellectual Development: Measuring Perry's Scheme." Journal of College Student Personnel, vol. 24, (1983), pp. 6-12.

Perry, W.G., Jr. Forms of Intellectual and Ethical Development in the College Years. New York: Holt, Rinehart and Winston, 1970.

Measure of Epistemological Reflection

Author: Margaret Baxter-Magolda, Department of Educational Leadership, Miami University, Miami, OH; Scales: Total Score; Length: 6 stimuli; Time: untimed.

The Measure of Epistemological Reflection (MER) represents a bridge between recognition and production measures. Six stimuli corresponding to Perry's levels of development are presented to subjects, who are then asked to justify the reasoning used in each stimulus. Standardized scoring procedures provide a quantified measure of intellectual development. Alpha reliability for the ratings may be as high as .76, while interrater reliability has ranged from .67 to .80, depending on the amount of training provided to raters. Research has provided support for the developmental underpinnings of the MER, revealing significant score differences for different educational levels.

Baxter-Magolda, M. and Porterfield, W.D. "A New Approach to Assess Intellectual Development on the Perry Scheme." Journal of College Student Personnel, vol. 26, (1985), pp. 343-351.

Reflective Judgment Interview

Authors: K.S. Kitchener, School of Education, University of Denver, Denver, CO, and P.M. King, Department of College Student Personnel, Bowling Green State University, Bowling Green, OH; Scales: Total Score; Length: four dilemmas; Time: approximately 40 minutes.

Like the MER, the Reflective Judgment Interview (RJI) represents a bridge between recognition and production measures. It consists of four dilemmas which are presented individually to the subject. Each dilemma is followed by a series of standardized questions designed to identify which of Perry's seven stages of intellectual development is being used by the subject to deal with that dilemma. A subject's score is the average rating across dilemmas and across raters. Research has shown that the RJI evidences acceptable levels of reliability (.73 to .78). In addition, the RJI has been found to be significantly related to other measures of critical thinking, as well as to levels of education.

Brabeck, M.M. "Critical Thinking Skills and Reflective Judgement Development: Redefining the Aims of Higher Education." Journal of Applied Developmental Psychology, vol. 4, (1983), pp. 23-34.

King, P.M. and Kitchener, K.S. "Reflective Judgment Theory and Research: Insights into the Process of Knowing in the College Years." Paper presented at the annual meeting of the American College Personnel Association, Boston, 1985. [ED#263-821]

Kitchener, K.S., and King, P.M. "Reflective Judgment: Concepts of Justification and Their Relationship to Age and Education." Journal of Applied Developmental Psychology, vol. 2, (1981), pp. 89-116.

Test of Thematic Analysis

Author: David G. Winter, Department of Psychology, Wesleyan University, Middletown, CT 06457; Scales: Total Score (optional: differentiation, discrimination, integration); Length: one exercise; Time: approximately 30 minutes.

The Test of Thematic Analysis uses a compare and contrast format to assess critical thinking skills. Subjects are presented with two sets of data and are asked to describe (in writing) how the two sets differ. The content of the essays is scored on a nine-point scale. In addition, scales derived from human information processing research can be used to evaluate the structure of the responses. Studies have found high levels of interrater agreement when scoring the TTA. Test scores also have been found to be significantly correlated with academic ability and coursework. In addition, measures of the structural characteristics of students' essays have been found to be significantly related to other measures of critical thinking, as well as to previous educational experiences.

Schroder, H.M., Driver, M.J. and Streufert, S. Human Information Processing. New York: Holt, Rinehart and Winston, 1967.

Winter, D.G. Thematic Analysis: An Empirically Derived Measure of Critical Thinking. Boston: McBer and Company, 1967.

Winter, D.G. and McClelland, D.C. "Thematic Analysis: An Empirically Derived Measure of the Effects of Liberal Arts Education." Journal of Educational Psychology, vol. 70, (1978), pp. 8-16.

Winter, D.G., McClelland, D.C. and Stewart, A.J. A New Case for the Liberal Arts. San Francisco: Jossey-Bass, 1981.

Watson-Glaser Critical Thinking Appraisal

Publisher: G. Watson and E.M. Glaser, Harcourt, Brace, and World; New York, NY; Scales: Total Score, Inference Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments; Length: 100 items; Time: 50 minutes.

The Watson-Glaser Critical Thinking Appraisal (CTA) is a multiple-choice measure designed to assess students' critical thinking abilities. In addition to a total score, five sub-scores can be derived from the CTA. Research has found that the total score on the CTA evidences acceptable reliability (.85 to .87) over seven norm groups and that students' performance on the CTA is positively related to their college experiences. In addition, the CTA has been found to be predictive of performance in courses emphasizing critical thinking.

Crites, J.O. "Test Review." Journal of Counseling Psychology, vol. 12, (1965), pp. 328-330.

Helmstadter, G.C. "Watson-Glaser Critical Thinking Appraisal." Journal of Educational Measurement, vol. 2, (1965), pp. 254-256.

Westbrook, B.W. and Sellers, J.R. "Critical Thinking, Intelligence, and Vocabulary." Educational and Psychological Measurement, vol. 27, (1967), pp. 443-446.

Wilson, D.G. and Wagner, E.E. "The Watson-Glaser Critical Thinking Appraisal as a Predictor of Performance in a Critical Thinking Course." Educational and Psychological Measurement, vol. 41, (1981), pp. 1319-1322.

Assessment of Values

Defining Issues Test

Author: James R. Rest, Department of Social, Psychological and Philosophical Foundations of Education, 330 Burton Hall, University of Minnesota, Minneapolis, MN 55455; Scales: "p" score; Length: 72 items; Time: untimed.

Rest developed the Defining Issues Test (DIT), a recognition measure of moral reasoning, based on the six stages identified by Kohlberg. Research has indicated that alpha reliability for the DIT is .77 and test-retest reliability is approximately .80. Research also has indicated that the DIT is significantly correlated with other measures of moral development, specifically Kohlberg's measure, and longitudinal research has found evidence of progression from lower-ordered to principled reasoning. Results also indicate that the DIT produces higher scores for principled reasoning than does Kohlberg's measure, and these higher scores are not due to upward faking on the DIT. These results suggest that production and recognition measures provide significantly different views of moral reasoning.

Biggs, D.A. and Barnett, R. "Moral Judgement Development of College Students." *Research in Higher Education*, vol. 14, (1981), pp. 91-102.

Davison, M.L. and Robbins, S. "The Reliability and Validity of Objective Indices of Moral Development." *Applied Psychological Measurement*, vol. 2, (1978), pp. 391-403.

McGeorge, C. "Susceptibility to Faking the Defining Issues Test of Moral Development." *Developmental Psychology*, vol. 11, (1975), p. 108.

Rest, J.R. "Longitudinal Study of the Defining Issues Test of Moral Judgement: A Strategy for Analyzing Developmental Change." *Developmental Psychology*, vol. 11, (1975), pp. 738-748.

Rest, J.R. *Development in Judging Moral Issues*. Minneapolis: University of Minnesota Press, 1979.

Rest, J.R., Cooper, D., Coder, R., Massanz, J. and Anderson, D. "Judging the Important Issues in Moral Dilemmas--An Objective Measure of Development." *Developmental Psychology*, vol. 10, (1974), pp. 491-501.

Humanitarian/Civic Involvement Values

Author: Ernest T. Pascarella, College of Education, University of Illinois at Chicago, Box 4348, Chicago, IL 60680; *Scales*: Total Score; *Length*: 6 items; *Time*: untimed.

The measure was derived from questions on the survey designed by the Cooperative Institutional Research Program (CIRP). Alpha reliability for this scale has been estimated to be .77. Results of research using this scale indicate that collegiate academic and social experiences are significantly related to the development of humanitarian/civic-involvement values, and that social involvement has the greater impact.

Pascarella, E.T., Ethington, C.A. and Smart, J.C. "The Influence of College on Humanitarian/Civic-Involvement Values." Paper presented at the annual meeting of the American Edcational Research Association, Washington, D.C., 1987.

Kohlberg's Measure of Moral Development

Author: Lawrence Kohlberg, "The Development of Modes of Moral Thinking and Choice in the Years Ten to Sixteen." Unpublished Doctoral Dissertation, University of Chicago, 1958; Scales: Total Score; Length: three dilemmas; Time: untimed.

In an effort to assess moral reasoning, Kohlberg developed a production measure that presents subjects with three moral dilemmas and requires them to explain how the dilemmas should be resolved. Subjects' responses are scored by raters trained to identify the dominant stage of moral reasoning employed. Reliability estimates for this technique are well within accepted limits (above .90). Research has provided support for the construct validity of Kohlberg's approach, identifying a clear step-by-step progression through the stages of moral reasoning. Moral reasoning also has been linked to students' previous educational experiences.

Kohlberg, L. The Psychology of Moral Development. New York: Harper and Row, 1984.

Rokeach Value Survey

Publisher: Halgren Tests, The Free Press, New York, NY; Scales: Instrumental Values, Terminal Values; Length: 36 items; untimed.

The Rokeach Value Survey was designed as a means of describing subjects' value systems. Respondents are asked to rank two sets of values (instrumental and terminal). Multiple administrations of the instrument can be used to measure stability and change in value systems. Test-retest reliability has been estimated to be adequate (.65 to .74). Moreover, research has shown that changes in individuals' value systems can be linked to life events.

Rokeach, M. The Nature of Human Values. New York: The Free Press, 1973.

Rokeach, M. (ed.) Understanding Human Values: Individual and Societal. New York: The Free Press, 1979.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1988— 2 1 7 -0 4 4 /0

APPENDIX H

Tables

TABLE 1: ORIGINATING STATES FOR REQUESTS

STATE	REQUEST FOR PUBLICATIONS	REQUEST FOR GENERAL INFO	TOTAL REQUESTS
Alaska		1	1
Arkansas	2	4	6
Alabama	5	33	38
Arizona	4	13	17
California	3	72	75
Colorado	3	62	65
Connecticut		14	14
Delaware	1	4	5
Florida	5	34	39
Georgia	3	33	36
Hawaii		15	15
Illinois	6	34	40
Indiana	11	31	42
Idaho	1	4	5
Iowa	2	15	17
Kentucky	8	34	42
Kansas	4	11	15
Louisiana	3	33	36
Maryland	2	18	20
Massachusetts	3	7	10
Missouri	10	47	57
Michigan	5	26	31
Minnesota	6	24	30
Mississippi	2	18	20
Maine	1	5	6
Nebraska	1	3	4
New Hampshire		5	5
New Mexico	1	12	13
New Jersey	4	61	65
New York	8	44	52
North Carolina	6	47	53
North Dakota		3	3
Nevada		3	3
Oklahoma		4	4
Oregon	1	7	8
Ohio	6	26	32
Pennsylvania	4	44	48
Rhode Island	1	14	15
South Carolina	6	26	32
South Dakota		3	3
Tennessee	8	73	81
Texas	5	51	56
Virginia	6	67	73
Vermont	1	4	5
Utah	1	2	3
Washington	1	9	10
Wisconsin	1	33	34
West Virginia	1	5	6

Wyoming	1	1	2
District of Columbia	3	52	55
Puerto Rico	2	9	11
Canada		5	5
Australia		2	2
Israel		1	1
Germany		1	1
Netherlands		2	2
Unknown		34	34
Total	158	1245	1403

TABLE 2: REQUESTS BY MONTH RECEIVED

MONTH	REQUEST FOR PUBLICATIONS	REQUEST FOR GENERAL INFO	TOTAL REQUESTS
October, 1986		10	10
November, 1986		4	4
December, 1986		5	5
January, 1987		13	13
February, 1987		22	22
March, 1987		29	29
April, 1987		17	17
May, 1987		57	57
June, 1987		13	13
July, 1987		20	20
August, 1987		33	33
September, 1987		16	16
October, 1987		15	15
November, 1987		37	37
December, 1987		55	55
January, 1988		25	25
February, 1988		18	18
March, 1988		25	25
April, 1988		27	27
May, 1988		25	25
June, 1988		29	29
July, 1988	3	23	26
August, 1988	12	41	53
September, 1988	21	37	58
October, 1988	20	57	77
November, 1988	11	30	41
December, 1988	10	11	21
January, 1989	10	1	11
February, 1989	7		7
March, 1989	4	8	12
April, 1989	4	12	16
May, 1989	12	5	17
June, 1989	8	13	21
July, 1989	21	13	34
August, 1989	5	7	12
September, 1989	10	3	13
Not Dated		487	487
Total	158	1243	1401

TABLE 3: REQUESTS BY INSTITUTION

INSTITUTION	REQUEST FOR PUBLICATIONS	REQUEST FOR GENERAL INFO	TOTAL REQUESTS
Universities	82	561	643
Colleges	46	263	309
Community Colleges	23	119	142
Vol/Tech Schools	3	10	13
Professional Schools	1	14	15
Government Agency	1	57	58
Private Concerns	1	17	18
Others		80	80
Unknown	1	122	123
Total	158	1243	1401

TABLE 4: QUALITY OF ARC PUBLICATIONS & SERVICES

PUBLICATION OR SERVICE	QUALITY RATING*	NUMBER OF RESPONDENTS
Telephone Conversation	4.45	31
Assessment Bibliography	4.49	55
Articles on Assessment at UTK	4.06	54
Satisfaction Surveys used at UTK	3.92	39
Bibliography of Assessment Instruments	4.27	48
ARC Workshop Materials	4.14	35
Material on Locally Developed Tests	3.48	25
Research Papers on Assessment	4.00	34
Conference Papers or Presentations	4.00	43

Average Quality Rating = 4.13

---

*Scale was a 5-point Likert Scale ranging from 1 = poor quality to 5 = excellent quality.

TABLE 5: RESULTS OF THE ARC QUALITY SURVEY

ITEM	QUALITY RATING*
1. The material answered my questions	3.90
2. Center staff were helpful and courteous	4.32
3. Material was clearly written	4.03
4. Copies received were complete & readable	4.04
5. The time it took to get the material was reasonable	3.97
6. The cost of the material was reasonable	4.20
7. Information sent was helpful in implementing program	3.76
8. Overall, the service provided was useful in meeting our needs.	4.03

Average item rating = 4.03

---

*Scale was a 5-point Likert Scale ranging from 1 = strongly disagree to 5 = strongly agree.

TABLE 6: WORKSHOP EVALUATIONS

QUESTION	MARCH 1987	NOVEMBER 1987	NOVEMBER 1988
In comparison with other similar workshops you have attended, how would you rate the overall quality of the assessment workshop?	7.7	7.7	7.0
How would you rate the overall effectiveness of the workshop in addressing the assessment needs/questions of your institution?	7.5	7.3	6.6
Location of the workshop	7.9	6.6	7.3
Meeting facilities	7.7	7.0	7.5
Information received prior to the workshop	6.6	5.3	6.5
Registration assistance	8.7	8.2	8.5
Organization of the program		7.7	7.5
Opportunities to learn from others		7.7	7.2
Workshop materials	7.7	8.0	8.0
Length of the workshop	7.9	7.8	7.5
Workshop size		8.1	7.0
Quality of the Consultant Panel Presentations	7.4	7.6	7.0
Effectiveness of the Panel	7.4	7.3	6.6
Content of the concurrent sessions	8.1	7.8	7.2
Presentation of the content in the concurrent sessions	8.1	8.4	7.3
Effectiveness of the concurrent sessions		7.1	6.6
Interactions with similar institutions	6.1	6.5	6.7
Interactions with consultants	7.3	7.3	7.2

TABLE 7 : Workshop Quality

August 1987

QUESTION	MEANS
1. A. In comparison with other similar workshops you have attended, how would you rate the overall quality of this workshop?	7.30
B. How would you rate the overall effectiveness of this workshop in addressing the assessment needs/questions of your institutions?	6.71
C. Please rate each of the following:	
1. Location of the workshop (Knoxville)	8.10
2. Meeting facilities	8.24
3. Information received prior to the workshop	7.66
4. Registration assistance	8.98
5. Organization of the program	7.61
6. Opportunities to learn from others (at UTK and other institutions)	7.76
7. Workshop materials	8.05
8. Length of the workshop	8.24
9. Workshop size	6.20

D. Comments on general workshop quality and adequacy in meeting your needs:

Participants generally rated the quality and relevance of the workshop as good. The information provided was perceived as thorough and well documented. Suggestions included smaller workshops of similar size colleges with more interaction (question/answer format). Additional requests included more information on resource allocation, cognitive style, assessment of non-cognitive areas, and a notebook containing outlines of all presentations.

EFFECTIVENESS OF THE PANEL

QUESTION	MEANS
2. A. How would you rate the effectiveness of the discussion sessions in addressing the assessment needs/questions of your institution?	6.72
B. Using the 10-point scale, how would you rate each of the individual presentations?	
1. "Using Outcomes Information to Assess the Accomplishment of Institutional Mission and Goals" (Homer S. Fisher & Trudy W. Banta)	7.90
2. "System Environment Success Factors" (F. Ramsey Valentine)	6.31
3. "Using Existing Campus Information" (John T. Hemmeter)	6.10
4. "Gathering Information Through Surveys" (William Lyons)	8.13
5. "Gathering and Using Test Data" (Edward L. Medford)	6.44
6. "Work of the Assessment Resource Center" (Gary R. Pike)	6.71
7. "Technical Considerations in Using Surveys" (Hemmeter, Lyons, & Scroggins)	7.20
8. "Technical Considerations in Using Test Data" (Edward L. Medford)	6.16
9. "Motivating Participants and Reporting Results" (Trudy W. Banta)	8.08
10. "Research Using Assessment Data" (Gary R. Pike)	7.80

APPENDIX I

Letters of Support

Bethany
Lutheran College
Office of the Registrar

734 Marsh Street / Mankato, Minnesota 56001 / (507) 625-2977

April 15, 1988

Trudy Banta
University of Tennessee-Knoxville
2046 Terrace Avenue
Knoxville, Tennessee 37996-3504

Dear Trudy:

I want to take this opportunity to thank you and Gary for all your efforts in coordinating and sponsoring our recent COMP workshop in Kansas City. The conversations with the ACT folks and within the entire group were interesting and for me very educational. Many questions about COMP were raised which I had never really confronted.

I don't know if we will be able to meet together as a group, but I hope we will be able to continue to communicate in the future. Thanks again!

Sincerely,

Ron Younge

Ronald J. Younge
Registrar

RJY/dg

Bemidji State University BEMIDJI, MINNESOTA 56601

HONORS COUNCIL
218-755-3984

Dear Gary Pike and Trudy Banta:

Thanks so much for arranging the COMP discussion in Kansas City. I found the opportunity to share experiences and ideas most valuable. The controversy over reliability was in particular quite enlightening.

Yours sincerely,

Michael Field

"An Equal Opportunity Employer"

APPENDIX J

Excerpt from Cross-National Assessment of Higher Education

CROSS-NATIONAL ASSESSMENT
OF HIGHER EDUCATION

John Harris
Samford University
October, 1989

Primary Audiences

This report is intended for two audiences:

1. Officials in the U.S. Department of Education (USDE) who may use it to develop a Request-for-Proposals.
2. The Study Group on Evaluation of Higher Education sponsored by the Programme on Institutional Management in Higher Education (IMHE) of the Organization for Economic Cooperation and Development (OECD), Paris.

U.S. Questions

There appear to be three cross-national assessment questions that interest Federal officials involved in policies dealing with American higher education.

1. How does the quality of American higher education compare with that of other technological-industrial nations i.e. Australia, Canada, England, France, Japan, Russia, Sweden, West Germany?
2. Are there advantages for U.S. colleges and universities as well as the U.S. economy in some type of linkage with the European Community Course Credit Transfer System of the EC's Erasmus Bureau? If there are, what type of linkage could be established?
3. What are the free trade and balance of payments implications for freer cross-national recognition of professional credentials?

If there are compelling advantages, what are the feasible first steps toward increased cross national recognition?

Background Documents

With this report, the author completes his exploration of the possibility of a cross-national comparison of higher education. This exploration began in 1988 with the author's appointment to the five person Study Group on Evaluation of Higher Education sponsored by the Program on Institutional Management of Higher Education of the OECD.

The author has explored the feasibility of a direct comparison of the academic achievement of American and European students through testing. The interest in the feasibility of such a comparison originated with the Director of the Fund for Improvement of Postsecondary Education (FISPE).

The author's exploration of this possibility is reflected in the following documents:

1. "Cross National Assessment of Student Achievement: Issues and Opportunities To Be Explored," First Draft: January 16, 1989; Second Draft: January 31, 1989.
2. "Addendum," (This "Addendum" to the prospectus identified in #1 above was drafted in Paris in the week of January 23, 1989. It describes 12 papers (not counting the first one--Rationale and Plan) to be written by various authors on different aspects of a cross-national comparison of student achievement. Authors were identified for nine of the twelve papers and eight were eventually written and submitted to the Study Group.
3. Eight Mini-Papers
   (1) Alan Wagner (CERI-OECD), "The Context for a Cross-National Assessment of Student Achievement."

(2) Urban Dahllof, Professor of Education (Uppsala University, Sweden), "The Feasibility Study of Cross-National Assessment of Student Achievement in Higher Education." (Information needed to determine causes for deficiencies signalled by indicators as part of a strategy for improvements).

(3) Op cit, "Special Methodology Mini-Paper on Secondary School Preparation."

(4) Fritz Dalichow (Erasmus Bureau-Brussels), "Method of comparing higher education degree programs in terms of program length or duration and subject controls."

(5) Trudy Banta (University of Tennessee), "Possibilities For Cross-National Comparisons." (Method of comparing content and levels of expectation on examinations used at the time of degree completion whether for assessment of individual student achievement or for program evaluation in different degree programs, i.e. chemistry, history, etc.)

(6) Charlotte Kuh (Educational Testing Service), "On the Use of the Graduate Record Examinations as an Indicator of Cross-National Achievement." (Exploration of how the GRE Subject Examinations or other similar examinations could be used to provide internal comparisons of student achievement in under-graduate degree programs, i.e. business administration, chemistry, history, elementary education, etc.)

(7) Richard L. Ferguson (American College Testing), "Using Certification And Licensure Examinations To Compare Cross National Levels of Performance In The Professions." (Method of comparing content and pass levels of examinations that admit candidates to professional practice in such fields as, medicine, mechanical engineering, accounting (sub-field of business administration), and elementary or secondary school teaching.)

(8) Roeland in't Veld (Erasmus University, Netherlands), "Selection by Multi-National Corporations." (Method of investigating if and how multi-national corporations observe and possibly use differences in degree program outputs in recruiting and hiring executives or technical experts in such fields as business administration, chemistry, and mechanical engineering.)

4. Memo to Mini-Paper Authors--"Status of Cross-National Assessment Possibility." This memo summarizes the thinking of the Study Group and of the author on how higher education may be compared cross-nationally. It does not deal with the economic-labor issue involved in cross national recognition of professional credentials and transfer of academic credit. The economic-labor issue arose after this June, 1989, memo was drafted.

5. "Significant U.S.A. Assessment Documents" is an annotated bibliography prepared by the author at the request of the Study Group. The author provided copies of the publications included in this bibliography to the IMHE staff at the second meeting of the Study Group in 1988 in Paris.

Objective

The objective of this proposed cross-national assessment of higher education is to provide American higher education policy makers information on:

1. How American higher education compares with that of other industrialized nations in terms of 1) secondary school preparation, 2) selectivity, 3) curricular content, 4) expected student achievements as evidenced by examination procedures, and 5) degree completion rates.
2. The opportunities that the development and implementation of the European Community's policies and procedures for cross-national transfer of academic credit and the cross-national recognition of professional credentials may present for boosting educational exports and free trade in professional and business services.

Focusing Considerations

A general study plan is proposed later in this report. It is predicated on the following "focusing considerations":

Preparation and Process. The original interest in cross national comparison was on students' tested performance. As indicated in the "background documents," international testing of university students does

DAVID LIPSCOMB UNIVERSITY

January 6, 1989

Dr. Trudy Banta
Learning Research Center
The University of Tennessee, Knoxville
Knoxville, TN 37996-4350

Dear Dr. Banta:

The discussion sessions you have convened on assessment have been very helpful to me. As you know, I have been professionally engaged in assessment theory and work for some time. The general conferences that are held are of little value to me professionally. In contrast, your sessions allow those of us who are deeply involved to explore theoretical and implementation issues that help us advance the state of the art. It is important, therefore, that your Center receive additional funds for these activities.

My personal view is that we need considerable discussion on how assessment data can be collected, arrayed, and used for continuous improvement of educational practice. Therefore, I would favor sessions with experts, perhaps from business or industry, who use numbers for the improvement of quality of services and products. Higher educators can learn a great deal from them.

Sincerely,

John Harris

John Harris

JH/bd

Nashville, Tennessee • 37204-1951
269-1000
TN WATS 1-800-342-2100 • US WATS 1-800-251-2054

3401 S. 39th Street
Milwaukee, WI
53215-4020
(414) 382-6000

Alverno College

OFFICE OF RESEARCH AND EVALUATION

January 12, 1989

Dr. Trudy Banta
Learning Research Center
1819 Andy Holt Avenue
The University of Tennessee-Knoxville
Knoxville, Tennessee 37996-4350

Dear Dr. Banta,

This letter is in support of your efforts to obtain FIPSE funds for some additional dissemination activities for your third year project. The goal for these funds is to support some continuing seminars for those of us working to resolve the emerging measurement problems in higher education assessment. We understand that you are requesting funds for seminars similar to those you planned for 1988-89 on assessment issues.

We are very much in support of your efforts to organize some additional meetings of a small but experienced group who are actively engaged in creating assessment practices. We are particularly interested in those practices related to establishing the validity of assessment processes and instruments used by institutions to validate student outcomes of college--and to establish the validity of curricula.

In our view, both assessment processes and instruments designed by faculty to assess and credential individual student learning, and those designed by an institution's assessment specialists to establish the validity of curricula and programs are in need of careful validation. These locally-designed instruments are critical to the growing assessment movement, since many institutions and their faculties find that off-the-shelf measures created by testing companies are not useful for the more individualized purposes of institutions.

We believe that the validity of an assessment process can only be determined in light of educational purposes and principles, and that therefore we need new ways to think about validity, ways that begin with and are driven by educational purposes, not by the traditional purposes of measurement alone. In the current educational reform movement, educational purposes and principles are being re-examined, as are approaches to assessment.

It is not surprising that those of us in the thick of this re-examination may need some particular strategies to help us confront the measurement issues that are emerging. We need an atmosphere of support where new ideas can be tried out and problems can be identified and solved.

It is the consensus of our seminar group that reviewing these issues together will ultimately lead to problem identification--which is one of the main outcomes of our meetings so far. As a group, we are particularly interested in continuing these seminars because we find that assessment practice is currently being created to resolve various problems, often out of the experience of those faculty and other assessment specialists who are dealing with quite a different set of problems than has been faced previously by the measurement and evaluation professional community. Once we can identify problems in these seminars as a group, we can identify those individuals who might consult with us to resolve these problems. We will also find out which problems are currently wide-open, where there are no consultants who can help, and where we will have to find solutions out of our combined experience and own experimentation.

For example, each participating group member has independently come to the realization that finding ways to validate instruments that use open ended responses leads us to consider issues that our measurement consultants agree lie in uncharted territory. It is unlikely that we would have come to this insight had we not had the opportunity to formulate our thinking as a group--to examine what issues we had in common, and then to jointly pursue some means to solving our problems. In addition, we are identifying for ourselves what the current barriers to solutions are across institutions so that we can decide which problems and their solutions are most critical, and where resolution is most likely to benefit the broadest group of institutions.

The group process has been enormously valuable. It gives each of us confidence that we are not alone. The process demands clear thinking from each of us, and forces us to find the nub of the problems we are confronting. The opportunity to clarify in dialogue with others who have some understanding of the issues involved has been helpful because they are engaged in the same enterprise on a day to day basis. The amount of time we have created to be together takes us beyond simplistic definitions or the complaint stage of problem finding. Because we have different institutional contexts, we have a chance to identify commonalities in experience, as well as to share strategies so that each does not have to reinvent the measurement wheel.

Frankly, the meetings with consultants have been most helpful for us at Alverno, because the meetings have forced us to clarify our own positions on why and where current measurement theory has not been able to provide textbook answers to validating instruments. To say that we need new measurement approaches to establishing the validity of instruments (e.g., where the student performance rather than the item is the unit of analysis; where an open ended, complex, interactive, sustained, dynamic response is the type of response rather than a recognition response; where responses are performance based) has meant that we have had to clarify the nature of the problems we are finding. To say that qualitative validation strategies must be applied and that we have to study what it means to establish the validity of complex expert judgment (e.g., where the assessment situation can change per student; where the criteria applied can change per student) means reconsideration of just how we present the problem to the measurement community.

We emphasize how important it is to meet together, to know and trust one another, so we are willing to admit to our most difficult problems. Often these are problems that the professional measurement community claims to have solved. For example, there is widespread agreement that establishing the validity of expert judgment means, in part, establishing high interjudge agreement. It takes a supportive atmosphere and the support of colleagues who can imagine the problem to help us say publicly "no, not in every case" and then to respond to the polite "here's the text from Measurement 101; just add up times they agree and get a percent agreement" with the challenge "that doesn't necessarily answer the question and I am not sure why." Such dialogue has helped us sort out problems where answers exist from those that have no answers. Engaging measurement specialists in the search for new solutions becomes an essential part of our seminar strategy.

Now for the practical side. Many of us in this group of seminar participants are from institutions who are the most experienced in assessment. We find it difficult to meet at those professional meetings we jointly attend. We are usually presenting at these meetings and using outside session time to meet with people who are pressing us for what we already know. Many of us are making presentations on assessment at many professional meetings and cannot stay for a whole meeting. Our schedules conflict and allow only a single lunch or dinner meeting. Further, our institutional support goes for travel to professional meetings, and phone, computer and mail costs to maintain relationships with each other after our "seminars." We cannot afford additional travel.

In sum, the issues are there, the problems are critical, and the support is essential. We are in strong support, Dr. Banta, of your request to acquire FIPSE funds to help us all continue the work we have really just begun. We are prepared to contribute the time away from our other pressing concerns to begin to deal seriously with the measurement issues confronting higher education assessment that can benefit a range of institutions. We hope that FIPSE can provide some financial support.

Sincerely,

Marcia Mentkowski

Marcia Mentkowski, PhD
Director
Office of Research and Evaluation
Professor of Psychology

Glen Roger

Glen Rogers, PhD
Research Associate
Office of Research and Evaluation

JMU

James Madison University
Office of Student Assessment

January 6, 1989

Dr. Trudy Banta
Learning Research Center
University of Tennessee
1819 Andy Holt Avenue
Knoxville, TN 37996-4350

Dear Trudy,

I am writing for support of the University of Tennessee under FIPSE sponsorship to extent the assessment director seminars. These seminars bring together the most active people in assessment from colleges and universities around the country. From my perspective, these seminars serve several purposes:

1. To bring together professionals from active programs to meet, share ideas, and build informal support networks,
2. To allow for in-depth discussion of some advanced topics usually stimulated by an outside consultant,
3. To allow for focused discussions on future directions for the assessment movement in higher education.

In addition, most of us in these seminars feel a responsibility to work with other institutions which are relatively new to assessment. For example, my institution was chosen by our state to serve as a model for other Virginia institutions. The ideas from professionals who are conducting more sophisticated research and practice also filters out to a broader audience beyond those who attend the seminars.

At my institution, no money is allocated for these types of seminars. The last seminar which I attended at Princeton was graciously funded by The University of Tennessee using FIPSE monies. I will not be able to attend the Milwaukee seminar due to lack of funds.

Harrisonburg, Virginia 22807
(703) 568-6706

I hope these seminars are able to continue and appreciate the role FIPSE has played in funding support for assessment around the country.

Sincerely,

T. Dary Erwin
Director

ks.027